More than just a feeling?
Using wellbeing-in-the-moment to advance mental wellbeing measurement and intervention

Amanda Jane Henwood
Date: August 2023
Student ID number: 201819102
Supervisor: Professor Paul Dolan

Introduction and Methods: 8531 words
Critical Discussion: 4593 words

Copyright statement

Acknowledgements

I dedicate this thesis to my mum, dad and sister. Your support and interest have never been lacking and am I incredibly lucky for that, amongst other things. The person closest to me, Bentley Amankwah, and my best friends Cansu Nefesoglu, Cristal Ewers, Lucy Seijo, Emily Mehmet and Robyn Lissimore also deserve a very special thanks. You have all supported and shaped this work, often in ways you likely don't realise, and I am very grateful for that.

I would like to express the deepest level of thanks to my supervisor and mentor throughout this project, Professor Paul Dolan, who has allowed me to share in his pearls of wisdom and learn from the very best. Your guidance, frequent challenging and all-round commitment to my academic progress has been invaluable.

Another person who has provided incredible support and guidance from start to finish of my PhD is Dr Aleksandar Matic. I thank you for your insights, innovations and ingenious ideas. I would like to extend this thanks to all the inspiring people at Koa Health who supported my PhD, with special thanks to Oliver Harrison for seeing potential in me and allowing this collaboration to happen.

I would like to thank all my co-authors Dr Dario Krpan, Professor Mike Rinck, Dr Luc Schneider, João Guerreiro and Dr Teodora Sandra Buda. I have learned a lot from each of you.

I owe another special thank you to Dr Sebastian Nastuta who has been instrumental in helping me to navigate my way through the wonderful world of R statistical software. Your patience and willingness to help me out with last minute coding errors is appreciated more than you know.

I thank all friends and colleagues who have sparked ideas for this PhD and offered me general support throughout the years, you know who are.

Abstract

Mental wellbeing is a priority for people and governments worldwide. Improvements in mental wellbeing are based on our ability to measure and influence it. How people feel in the moment (their wellbeing-in-moment – WIM) is strongly related to mental wellbeing. Given its strong association with mental wellbeing and its malleability, WIM presents a promising avenue through which to better measure and influence mental wellbeing. The four papers presented in this thesis draw upon a range of methodological approaches to probe the interrelationship between WIM and mental wellbeing. Papers 1 and 2 explore how WIM relates to mental wellbeing to improve the conceptualisation and measurement of these constructs. Papers 3 and 4 explore the potential of two WIM-related processes/behaviours to improve mental wellbeing. Overall, this work highlights the potential of moments in time to advance psychological and behavioural science by informing mental wellbeing 1) measurement and, 2) intervention. The thesis concludes with the presentation of a novel framework that facilitates the integration and synchronisation of WIM-based mental wellbeing intervention research to increase its impact.

Table of Contents

1. Introduction 6

1.1. Definitions 8

1.1.1. Wellbeing-in-the-moment..8

1.1.2. Subjective wellbeing...11

1.1.3. Mental illness...13

1.1.4. Overall mental wellbeing..14

1.2. Topic refinement and focus 15

1.3. Key papers 17

1.3.1. Measure...17

1.3.2. Intervene 22

1.4. A framework for guiding and synthesising WIM-related mental wellbeing research 27

1.5. Summary of key papers 28

2. Methods……………………………………………………………29

2.1. Paper 1 "The duration of daily activities has no impact on measures of overall wellbeing"... 29

2.2. Paper 2 "Blurred lines? On the distinction between WIM and mental illness. A two-study replication" 31

2.3. Paper 3 "Pandemic-related changes in social interaction are associated with changes in automatic approach-avoidance behaviour" 32

2.4. Paper 4 "Reporting on wellbeing-in-the-moment, thoughts and context reduces anxiety" 34

3. Empirical Work 37

3.1. Paper 1 38

3.2. Paper 2 58

3.3. Paper 3 101

3.4. Paper 4 136

4. Critical discussion 171

4.1. Measure: overarching contributions of papers 1 and 2............................... 172

4.2. Intervene: overarching contributions of papers 3 and 4 175

4.3. The WIM Intervention Framework 177

5. Introduction references 187

6. Critical discussion references 193

7. Paper 1, References 196

8. Paper 2, References 198

9. Paper 3, References 203

10. Paper 4, References 209

11. Paper 1, Appendix 214

12. Paper 2, Appendix 232

13. Paper 3, Appendix 243

14. Paper 4, Appendix 259

1. Introduction

Feeling good overall is judged by many to be the most important dimension of life (Adler et al., 2022). Approximations of good feelings overall, such as subjective wellbeing (SWB), are increasingly used to guide economic and social policy (De Neve et al., 2020). They have also been adopted by many nations as key markers of societal progress (Dolan et al., 2022). How bad people feel overall is also important. Maladaptive bad feelings, otherwise known as mental illness, constitute a considerable health burden; they cost the UK at least £117.9 million annually, which is around 5% of the UK's GDP (LSE News, 2022). There is a strong social and economic case, therefore, for improving our understanding of, and ability to influence, SWB and mental illness.

One key, yet often largely overlooked factor, with the potential to affect both SWB and mental illness is how people feel in the moment: their wellbeing-in-the-moment (WIM). WIM is both impactful and pervasive since it colours our everyday experiences and has a marked impact on our cognition and behaviour. For example, positive emotional states make us more attentive to positive features in the environment and negative emotional states make us more attentive to negative features in the environment (Schwarz, 2000, p.435). Feeling sad often makes us more systematic and rational in our behaviour, whereas feeling happy tends to make us more likely to rely on mental short-cuts and heuristics (Hertel et al., 2000).

Whilst WIM is closely linked to SWB and mental illness, it is more malleable and more tangible. Momentary emotional states can be easily manipulated as has been shown by the mood induction literature (see Joseph et al., 2020 for recent review). These manipulations have been shown to be reflective of everyday moods and often cause associated changes in biological, behavioural, and attentional processes (Lench et al., 2011). The immediacy of momentary emotional states also makes them more tangible since they are less subject to memory recall abilities than SWB and mental illness. We feel and witness them as they are being measured. For all these reasons, WIM presents an exciting avenue that may be used to help understand, and influence, SWB and mental illness.

This thesis utilises the power of WIM to offer two main contributions to psychological science. The first main contribution is to use WIM to improve the measurement of subjective wellbeing and mental illness. Understanding how and why different constructs are associated is paramount to developing accurate conceptualisations and measurements of those constructs (Flake et al., 2017). There is currently a dearth of research combining insights from WIM, SWB and mental illness related literature despite their strong associations with each other (Keyes, 2007; Csikszentmihalyi & Larson, 2014; Wood, 2010). Clear conceptualisation and effective measurement are the critical foundations upon which scientific advancement, and effective intervention, must take place (Psychological Science, 2018). Papers 1 and 2 of this thesis target these gaps in understanding. Paper 1 examines the usefulness of measuring WIM over time to approximate SWB by comparing two of the most popular WIM-based SWB measures. Paper 2 assesses the extent of potential overlap between WIM and mental illness constructs by testing whether WIM inductions impact mental illness reports.

The second main contribution of this thesis is to explore whether WIM-related processes/behaviours can be used to influence SWB and mental illness. Mental wellbeing intervention is key for optimising societal wellbeing and bringing people out of misery (Krekel, 2021). A better understanding of how WIM or WIM-related processes/behaviours can be used to influence SWB and mental illness, can pave the way for novel mental wellbeing interventions. Papers 3 and 4 of this thesis test two different WIM-related pathways to mental wellbeing. Paper 3 explores whether naturally occurring changes in WIM-related behaviour (social interaction) are associated with changes in automatic approach-avoidance tendencies (known to determine SWB and mental illness). Paper 4 assesses whether WIM reporting via SWB measurement can be used to influence SWB and mental illness.

Before the presentation of the 4 papers, the following 'Definitions' (Section 1.1). will provide definitions of WIM, SWB, mental illness and overall mental wellbeing. It will also discuss current theory on the relationships between WIM with both outcomes. Section 1.2.1 'Measure' will explain how papers 1 and 2 seek to improve our understanding of the interrelationship between WIM, SWB and mental illness to inform the conceptualisation and measurement of these constructs. Section 1.2.3 'Intervene' will explain how papers 3 and 4 aim to improve our understanding of how to influence SWB and mental illness by exploring WIM-related pathways to SWB and mental illness.

Section 2 'Methods' provides methodological context for each paper including the main research questions, an explanation of how the research questions were scoped and narrowed down, and an overview of how WIM, SWB and mental illness were measured in each study.

Section 3 'Empirical Work' presents full final drafts of the 4 papers that make up this thesis.

Section 4 'Critical Discussion' presents a combined assessment of the paper contributions to the field of psychological and behavioural science. It concludes by introducing a novel framework that can be used by researchers and practitioners interested in taking this research forward.

1.1. Definitions

1.1.1. Wellbeing-in-the-moment

The term wellbeing-in-the-moment is used throughout this thesis to describe momentary states pertaining to emotions, mood and affect. Emotions have been defined as mental appraisals of situations that comprise visceral, cognitive, and motor components (Cabanac, 2002). For example, 'fear' is a strong emotion that is both neurological and physiological (Steimer, 2022). It is associated with responses such as jumping or freezing and is restricted in duration to particular situations (Butler et al., 2007). Both mood and affect are on-going states that last longer than emotions and are often experienced with lower intensity (Gross, 2010). Emotions are generally thought to serve an immediate, adaptive purpose (James, 1884; Lewis, et al., 2008), whereas mood and affect are broader constructs that evolve and develop over time (Clark, 2005). Emotions can also give rise to moods and affective states and vice versa. For example, somebody in a bad mood might be more susceptible to emotions such as anger and resentment than somebody in a good mood (Clark, 2005). Similarly, somebody who experiences fear (emotion) may also report feeling nervous and apprehensive (mood) (Izard, 1977). WIM is a term used throughout this thesis that encompasses emotion, mood and affect. These states have been combined since the boundaries between them are often blurred and difficult to capture.

Traditionally, different types of emotions have been considered as separate constructs (e.g., Tomkins, 1962; Izard, 1972). This perspective is supported by a wealth of cross-cultural evidence highlighting the relationship of different emotions (e.g., fear and happiness) to facial expressions and biological responses (Ekman, 1972). The concept here is that different emotions have different outcomes that are distinct to their discrete categories. As a result, many researchers adopt discrete emotions (e.g., "how happy/sad/anxious do you feel right now?") to measure momentary emotion. These questions have the advantage of being straight forward and intuitive to answer.

However, there is increasing suggestion that this standpoint promotes an oversimplified view of emotion, whereby important factors, such as the interrelationship between different emotions and the cognitive appraisal by which emotions are accompanied, are not properly accounted for (Barrett, 2009). According to this viewpoint, emotions are closely interrelated and can be grouped according to their position along two or more dimensions, starting with valence and arousal. Perhaps the most popular and widely used theoretical model supporting this viewpoint is James Russel's (1980) Circumplex Model of Affect (see Yik et al., 2011 for a more recent development of this model). According to this model, emotions can be plotted on a two-dimensional circular space in which the vertical axis represents arousal and the horizontal axis represents valence, see Figure 1. Valence captures the extent to which an individual feels positive or negative and arousal captures the extent to which an individual feels activated (Kron et al., 2015). These two types of WIM combine to produce different kinds of emotional states or affective responses. For example, an individual might feel excited (positive valence/high arousal) or angry (negative valence/high arousal).

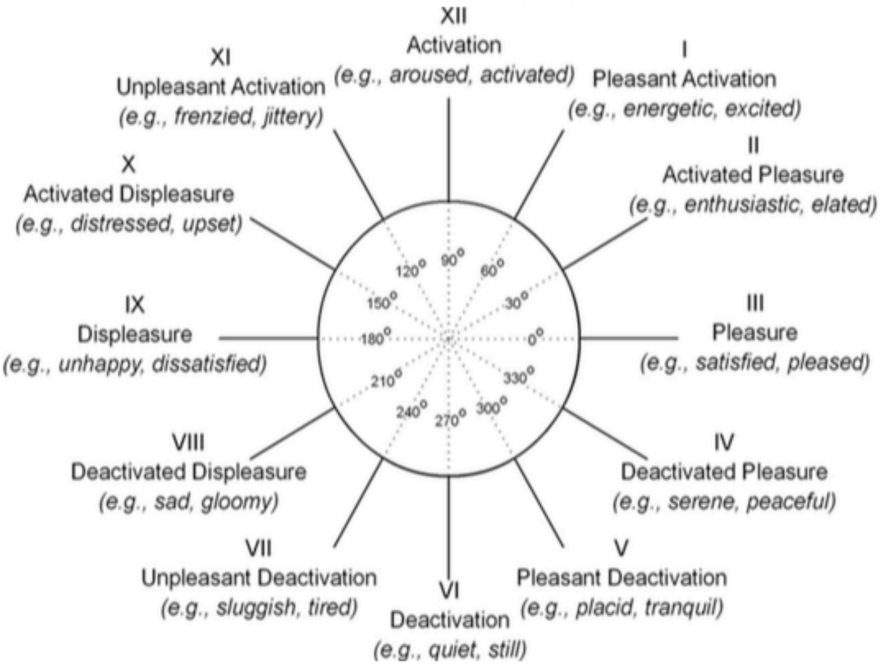


Figure 1. *A 12-Point Affect Circumplex. (12-PAC). This schematic diagram shows the hypothetical locations of the 12 segments of Core Affect.*

The papers in this thesis consider different approaches to capturing WIM depending on the context of the research. For example, since paper 1 compares existing of WIM-based SWB measures which capture WIM by asking people how happy they feel on a scale of 0-10, valence and arousal were not considered as the primary measurement approach. In addition to discrete emotions, there are also complex emotions, which typically represent a combination of basic emotions (Power & Tarsia, 2007). One example of a complex emotion is jealousy since this emotion can involve fear of loss, inadequacy, insecurity, and envy. Valence and arousal can capture the affective sentiment associated with both discrete and complex emotions and is therefore used in this thesis whenever appropriate. Explanations and justifications for the WIM measures used in each study can be found in Methods Section 2.

A novel addition to the affective literature is the inclusion of momentary states of purpose in paper 1 – that is – how meaningful something feels in the moment. Whilst arousal and valence are widely recognised as key components of affect, purpose is not. This purpose dimension was adopted from conceptualisations of experiential SWB which not only include how happy people feel overall but also how meaningful or purposeful people feel overall (Dolan, 2012).

1.1.2. Subjective wellbeing

There have been many conceptualisations of SWB. These can be broadly divided into two categories: evaluations and experiences (Dolan & Kudrna, 2016). Evaluative SWB refers to how somebody feels about their life overall and is typically measured at one specific time point. Experiential SWB refers to how somebody feels in the moment as they are experiencing their lives and is typically comprised of SWB reports aggregated over multiple time points. The former can be thought of as a more general cognitive appraisal of long-term affect whilst the latter is more concerned with actual fluctuations in momentary affect over time. The dimension of how purposeful something feels in the moment has also been highlighted as an important component of SWB (Dolan, 2015). From this perspective, overall wellbeing is not only determined by how much pleasure someone feels overall but also how meaningful or purposeful the things they do in their life feel.

For evaluative measures of SWB, people are often asked how satisfied they are with their life overall, or how satisfied they are with respect to certain domains, such as work or relationships (Dolan et al., 2008). To capture purpose, they may also be asked how meaningful their lives feel, or whether they consider their lives to be worthwhile (Dolan et al., 2017). For experiential measures of SWB, there are two core methods of measurement: Ecological Momentary Assessment (EMA) – or Experience Sampling Method (ESM) - and the Day Reconstruction Method (DRM). EMA and ESM ask people "how happy do you feel right now" in the moment at random intervals as they go about their daily lives (Stone et al., 1999). DRM, on the other hand, asks people to fill out a diary of activities from the previous day and report how happy they felt during each of those activities (Kahneman et al., 2004). There is some debate in the literature as to which of these experiential SWB measures ought to be preferred (Dolan et al., 2017).

For experiential classifications of SWB, there is some overlap between WIM and SWB given that WIM forms an integral part of SWB. Insofar as these classifications are concerned there is no need to prove a distinction between the two constructs since the conceptual distinction is clear: experiential SWB is as a collection of WIM reports, whereas WIM is one affective moment in isolation. There is necessary empirical overlap between these constructs. It has been necessary, however, to empirically distinguish evaluative SWB from WIM, since these are conceptualised as being distinct (Chamberlain, 1988). Since evaluative SWB is conceptualised as a relatively stable, trait based cognitive measure that reflects a person's overall evaluation of the quality of his or her life it ought not be influenced by transient affective experiences, i.e. WIM. Recent work conducting multiple studies exploring this overlap potential has largely supported their distinction, demonstrating relatively little overlap between WIM and evaluative SWB (life satisfaction) (Yap et al., 2017).

1.1.3. Mental illness

In addition to SWB, it is important to study affective outcomes that are associated with psychopathology by including the study of mental illness. This can help to provide a more complete picture of affective impact. For example, mental illness measures can be used to identify the most miserable members in society, and thus serve to highlight opportunities where misery may be most significant and maladaptive. It is the maladaptive component of mental illness that makes it important to study in addition to SWB. Mental illness can be defined as the ongoing experience of negative mental states that are associated with clinical diagnoses, such as worry and rumination (Westerhof & Keyes, 2010). It is the opposite to mental health which has long been described as the absence of mental illness. Mental illness measures typically take the form of questionnaires that ask people the extent to which they have been experiencing mental illness symptomology (recently or in general). An example mental illness item taken from the Trait section of the State Trait Anxiety Scale is “I wish I could be as happy as other seem to be”.

Whilst mental illness is conceptually distinct from SWB it is also increasingly associated with the absence of SWB (Keyes, 2013). However, a growing body of work suggests mental illness, whilst strongly related to SWB, should be considered as independent from it. A recent review combining over 80 peer-reviewed papers concluded that indicators of subjective wellbeing can occur with or without the presence of mental illness diagnoses (Iasiello et al., 2020). Both overarching constructs contain common but also differential antecedents that ought to be separately assessed. Interventions set out to improve overall feelings may influence SWB, mental illness, or both (van Agteren et al., 2021). Previous research has shown that SWB is correlated with but factorially distinct from measures that capture common mental illness symptomology (Headey et al., 1993; Keyes, 2005). It is important therefore to measure both SWB and mental illness in tandem when attempting to detect the impact of a given psychological intervention.

Nevertheless, much of SWB research fails to consider mental illness outcomes, and vice versa (van Agteren, 2021). This may be because despite their noted differences many consider SWB and mental illness to be somewhat interchangeable given their high correlation (Flèche & Layard, 2017). It may also be because one is considered to take priority over the other and researchers hold different views about which is superior, or because of differing methodologies and research aims applied in each of the two areas. This thesis takes the stance that since they are highly associated but also dissociable, considering these two constructs in tandem can make a greater contribution to the field overall. Considering one without the other misses an important opportunity to understand how and when they both interact to determine the broader outcome of affective improvements.

This thesis focuses on anxiety and depression mental illness measures since they are the most common mental illnesses and therefore have a high prevalence in the population (Chalder et al., 2012). In addition, they are disorders of mood, which means they may be more likely than other mental illness questionnaires to be influenced by WIM. Depending on the context of the research in question this thesis will rely on different types of SWB assessment: evaluations, experiences, and mental health symptomology. Each measurement will be justified within the context of the individual research.

1.1.4. Overall mental wellbeing

Throughout this thesis I use the term mental wellbeing to refer to the combination of SWB and mental illness as per van Agteren (2021). From this standpoint, SWB and mental illness are conceptualised as being two distinct, but largely related higher order affective constructs, that together form the basis of people's overall mental wellbeing. WIM, on the other hand, is conceptualised here as a lower order, momentary, construct that forms a part of experiential SWB and feeds into, but is mostly distinct from, evaluative SWB and mental illness. For robustness this assumption is tested directly where empirical work on the distinction is lacking (e.g., paper 2 tests the extent to which WIM is distinct from mental illness since there is a dearth of research detailing the extent to which these constructs overlap).

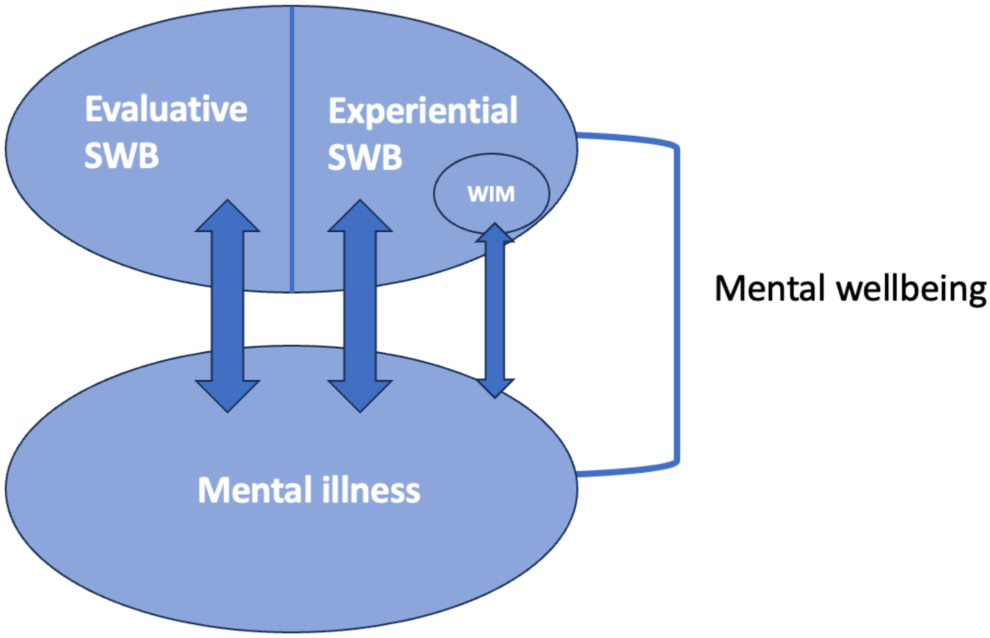


Figure 2. *Schematic diagram depicting the assumed relationships between WIM, SWB, mental illness and mental wellbeing. SWB can be conceptualised as either evaluative or experiential SWB. WIM is a momentary measure of wellbeing that forms an important part of experiential SWB, which is a collection of WIM experiences over time.*

1.2. Topic refinement and focus

There is a burgeoning literature on WIM, SWB and mental illness. It has been necessary to limit the scope of that literature within the current thesis. I began this process by splitting my papers into two pre-defined sections: **measure** and **intervene**. These areas of exploration were chosen since being able to influence how people feel is key to making a positive impact in the world and to be able to do so, a solid understanding of how to measure that impact is required. Therefore, the two areas provide important contributions that complement one another.

I narrowed the measure papers down by focusing on each of the two core constructs that make up mental wellbeing in turn: SWB and mental illness. Paper 1 focused on advancing SWB measurement by probing the validity of two of the most popular SWB measures, whilst paper 2 focused on advancing mental illness (anxiety and depression) measurement and conceptualisation. Both papers drew upon WIM to achieve these aims. The SWB measures in paper 1 used repeated reports of WIM to approximate SWB, and paper 2 explored whether mental illness measures contain variance associated with WIM.

I narrowed down the intervene papers (3 and 4) by focusing on key intervention components as opposed to full interventions. There are many different pathways to improving overall mental wellbeing most of which consist of cleverly crafted interventions that combine several different processes to produce an effect. However, intervention research can be better advanced once key processes that lead to changes in mental wellbeing have been isolated. This is because component isolation provides critical information about the mechanisms through which treatment may operate (Kazdin, 2007). As such, the present work was limited to a focus on isolated WIM-related behaviours (social interaction for paper 3 and WIM reporting for paper 4) as key intervention components. WIM-related behaviours are defined here as behaviours that have a direct association with WIM.

Social interaction was chosen since it is inherently associated with positive feelings (Ross & Inagaki, 2022) and the significance of this behaviour became particularly pertinent during the global pandemic context within which this paper was conducted. Insights about the impact of social interaction on overall mental wellbeing may also be usefully applied to many other contexts in which social interaction is likely to be affected. WIM reporting was selected as a possible intervention component since, given its simplicity, it holds promise as an easy and cost-effective mental wellbeing intervention.

1.3. Key papers

As described above the thesis is split into two parts: measure and intervene. Papers 1 and 2 form the basis of the measure section and papers 3 and 4 form the basis of the intervene section. Together, all papers further our understanding of the interrelationship between WIM, SWB and mental illness. The text below details the main contributions of each section and the papers within them.

1.3.1. Measure

Papers 1 and 2 address the first main contribution of this thesis: using WIM to inform the measurement of mental wellbeing. In general, they argue that mental wellbeing measurement may be improved by exploring the usefulness of WIM to 1) approximate SWB (paper 1) and, 2) probe our understanding of mental illness measurement (paper 2). These papers are detailed in turn below.

Paper 1 " **The duration of daily activities has no impact on measures of overall wellbeing"** explores the research questions:

- *Does the duration of affective moments impact overall (experiential) SWB scores?*
- *Do Ecological Momentary Assessment and Day Reconstruction Method produce comparable SWB reports?*

WIM as a measure (Experiential SWB)

There has been considerable debate over which types of SWB we should optimise for. Economists have tended to prioritise evaluative SWB measures such as life satisfaction because they are easy to answer and frequently included in international surveys

(Dolan et al., 2011). Evaluative measures such as life satisfaction and general happiness have also been correlated with factors we might assume to be related to SWB such as absolute and relative income, employment status, relationships, health, major life events and personal characteristics such as age, gender, and personality (see review by Dolan et al., 2008; Waldron, 2010). Thus, evaluative SWB appears to have been useful for establishing an overall sense of how people are doing.

However, evaluations of how well people are doing overall can fail to capture important nuance about how people experience their lives day-to-day which can be captured by momentary measures. This is because someone asked to indicate "overall, how satisfied are you with your life" might focus on one or more key factors when answering this question, such as whether they have a good job or how they are feeling that day, extrapolating how they feel about their lives overall from that focal point. This "focusing effect" has been well established empirically in the context of SWB as well as other domains (Dolan & Powdthavee, 2012; Dolan & Metcalfe, 2010; Kahneman et al., 2006). When forced to estimate how they are feeling about their lives overall, it is also likely that individuals may be unduly influenced by social norms surrounding which factors *should* lead them to be satisfied with their lives, such as having had children by a certain age or holding down a job that society defines as successful (Dolan, 2015).

As well as helping to avoid focusing effects, momentary measures also consider time use which can reveal important insights about how to optimise SWB. For example, research comparing evaluative with experiential SWB measures has shown that despite being dissatisfied with their lives overall, unemployed people report higher SWB during their daily experiences relative to employed people (Knabe et al., 2010). This is explained by the fact that they use their time in more enjoyable ways than their employed counterparts do. Therefore, SWB measures that account for how people experience their lives moment-to-moment appear crucial to achieving a full picture of

SWB since they capture aspects of everyday thriving that might otherwise be omitted by global, evaluative measures.

Given these advantages, debates surrounding SWB measurement have placed increasing emphasis on the importance of conceptualising SWB as a collection of experienced moments as opposed to one shot global assessments (Dolan, 2015; Kahneman et al., 2004). However, our ability to prioritise the significance of moments over time, is reliant on effective and efficient measurement (Dolan et al., 2011; Griffin, 1986). Within this context, paper 1 of this thesis sought to explore the effectiveness and efficiency of two of the most popular WIM-based SWB measures: Ecological Momentary Assessment and the Day Reconstruction Method.

This paper explored the effectiveness of EMA and DRMs by testing the current assumption that these measures produce consistent measures of experiential SWB. This is important since if both measures produce consistent SWB scores we can be more confident in our assumption that they are picking up on the same underlying construct of experiential SWB. Previous research supporting this assumption has been limited by its reliance on collecting SWB reports from small, non-diverse samples over short time periods of just 1-2 days (Kahneman et al., 2004; Dockray et al., 2010). Paper 1 sought to solidify these findings by testing the consistency of the most popular experiential SWB measures on a much larger and more diverse sample over a longer time frame of 2-3 weeks.

The efficiency of current measures was also assessed by testing whether EMA and DRM capture duration in a way that is meaningful for overall SWB. Duration is currently considered to be a necessary, though time consuming, addition to SWB reports. However, thus far no research has clarified whether including additional questions about how long people spend in the affective episodes captured by EMA and DRMs is contributing anything to the approximation of SWB. If duration does not change SWB scores then this would raise important questions about the validity of this measure, since longer durations spent in happy moments ought to contribute to higher SWB. By testing these core assumptions upon which WIM-based SWB measurement is based, this research presents a significant methodological contribution to the literature of how best to optimise WIM measurement over time to generate a more robust measure of experiential SWB.

Paper 2 " **Blurred lines? On the distinction between WIM and mental illness. A two-study replication"** explores the research question:

- *Does manipulating WIM change anxiety and depression reports?*

Moments as determinants of a measure (mental illness)

In addition to forming the basis of SWB measurement, WIM can also be used to better understand the extent to which measures of SWB and mental illness relate to WIM. Whilst WIM has already been successfully distinguished from evaluative SWB (and forms an important part of experiential SWB), less is known about its relationship with mental illness. Assessing the relationship between WIM and mental illness can add important clarity to this distinction.

WIM may impact on mental illness reports in two important ways. First, it is possible that how people feel in the moment (WIM) might change the way they respond to questions about their mental illness. According to the Associative Network Theory, negative moods make negatively valanced information more salient whilst positive moods make positively valanced information more salient (Bower, 1981). This is because they activate a network of similar associations in the brain. If somebody is experiencing a negative emotion, therefore, they might be more likely to recall times in which they felt negative in the past, thus biasing their response to items that ask about the frequency at which they experience negative emotional states.

Second, is also possible that the wording of mental health questionnaires may interact with WIM to prime responses in ways that promote category confusion. Specifically, the wording of mental illness questionnaires may prime individuals to interpret their underlying emotional states in different ways. For example, somebody in a state of negative high arousal might feel anxious when looking at a series of questions in which anxiety is the frame, e.g. “how anxious do you feel?”. Yet, if exposed to a series of questions in which depression is the frame, e.g. “how sad do you feel?” they might just as easily be primed in the opposite direction, interpreting their emotions as synched with those of depressive symptomology. Crucially, this means that the same underlying emotion may be interpreted by the same person in different ways depending on the context within which that individual finds themselves.

To this end, paper 2 tests whether manipulating WIM impacts upon two of the most prevalent mental illness measures: anxiety and depression. This paper includes two online studies in which participants were asked to complete an autobiographical memory recall WIM induction task designed to bring about specific emotional states: happiness (high arousal high valence), sadness (low arousal low valence), anxiousness (high arousal low valence) and relaxation (low arousal high valence), or a neutral mood induction. Following WIM induction, participants then complete anxiety and depression questionnaires presented in a randomised order to see whether WIM condition impacts scores.

Assessing the relationship between WIM and mental illness is important for several reasons. First, it can help to affirm the validity of current mental illness measures. A measure is generally considered to be effective if it is not distorted by subtle contextual factors, such as how people feel (Schwarz & Strack, 1999). Second, it can inform theoretical understandings of the relationship between WIM and mental illness. Mental illness is increasingly conceptualised as being on a spectrum with WIM, whereby anxious and sad moods lie at one end and severe anxiety and depression lie at the other (Chentsova-Dutton et al., 2015). If true, then one might expect WIM induction to influence mental illness reports since mental illness is simply a more intense and maladaptive form of WIM, thus the two share overlapping WIM related variance. WIM induction should impact anxiety and depression reports since the two must presumably overlap at some point. Third, clarifying the distinction (or lack of) between these constructs can help to inform our understanding of mental illness interventions. Importantly, if these two constructs are overlapping then any identified impact of WIM-related interventions on mental illness might be detected, not because WIM is influencing mental illness, but because WIM has changed and the mental illness measure is detecting that change. Since it may be possible to impact the negative WIM associated with mental illness whilst not changing the underlying tendency itself, distinguishing between these impacts is of practical significance.

1.3.2. Intervene

Papers 3 and 4 address the second main contribution of this thesis: to highlight WIM-related pathways to SWB and mental illness intervention. The following text explains the rationale behind why a focus on SWB *and* mental illness, as well as a focus on WIM, is important for the advancement of psychological interventions. This text is followed by an overview of paper 3 and 4 in turn.

Many people remain in perpetual negative affective experiences. As a result, identifying factors that allow us to intervene and alleviate such experiences is crucial if we want to improve the lives of the most miserable in our society. Mental wellbeing interventions can be defined as activities aimed at changing behaviours, feelings, and emotional states (Hodges et al., 2011). They have typically been split into positive psychological interventions (PPIs) that target subjective wellbeing, and clinical interventions, that target the alleviation of mental illness. PPIs tend to target increased positive feelings, thoughts, and behaviours whereas clinical interventions tend to target maladaptive behaviours and cognitive processes such as worry and rumination (van Agteren et al., 2021).

As a result of this segregation, there remains a considerable gap in the literature with respect to which kinds of interventions promote changes across which mental wellbeing dimensions (e.g., just subjective wellbeing, just mental illness, or both). A more extensive mapping of influence factors onto affective outcomes can help to build a clearer picture of the interrelationship between SWB, mental illness and the relevant factor of influence. For example, if we know that increasing social behaviour increases wellbeing *and* alleviates mental illness then we can assume that SWB and mental illness are both related to this shared behaviour. Over time, a more complete mapping of behavioural processes onto these two key constructs can build a comprehensive picture of how different kinds of interventions relate to SWB and mental illness. Mapping of this kind can help to determine which kinds of interventions are best suited to which people. For example, an individual with low levels of SWB but also low mental illness symptomology can be given an intervention that specifically targets the behaviours associated with SWB and not mental illness. Whereas an individual with both low SWB and high mental illness symptomology may be better placed to complete behavioural interventions known to target both.

Whether and how WIM can be used to influence overall mental wellbeing outcomes is a factor that has been largely absent from current understandings of mental wellbeing interventions. Since WIM is strongly related to both SWB and mental illness, but is more immediate and malleable, it has potential as a pathway through which to influence SWB and mental illness outcomes. The importance of WIM for inducing psychological change conducive to improvements in mental wellbeing has long been postulated in clinical psychology (Lane et al., 2015). For example, Freud fixated on the importance of emotional expression during trauma to mitigate its detrimental effects (Freud, 1899-1999). More recent theorists have stressed the importance of in-session arousal during psychotherapy to help patients develop a motivation to experience and overcome uncomfortable emotions (Aafjes, 2017). Yet despite the theoretical attention it has received, few studies have systematically investigated the role of WIM in facilitating mental wellbeing enhancing interventions (McCullough et al., 2011). This is critical since even though there are now numerous interventions that target improvements in mental wellbeing, we still lack a definitive understanding of how they work (Kadzin, 2009).

*Paper 3 "**Pandemic-related changes in social interaction are associated with changes in automatic approach-avoidance behaviour"*** explores the research question:

- *Are changes in WIM-related behaviours (social interaction behaviours) associated with changes in mental wellbeing related outcomes (approach-avoidance tendencies)?*

To highlight potential WIM-related pathways to SWB and mental illness intervention, paper 3 explores the association between changes in social interaction with changes in automatic approach-avoidance behaviour over time in a COVID-19 context.

The WIM-related behaviour in focus here is social interaction. Social isolation is assumed to be an inherently negative experience given its adverse consequences for mental illness (Ganesan et al., 2021; Loades et al., 2020; Pancani et al., 2021; Ganesan et al., 2011). The COVID-19 pandemic provided a unique opportunity to test whether reducing social interaction was associated with automatic processes known to promote mental wellbeing: approach-avoidance tendencies. Therefore, the mental wellbeing related processes of focus were approach-avoidance tendencies. There is an important gap in the literature with respect to how real-world WIM-related behaviours impact upon approach-avoidance, despite much evidence highlighting this relationship in reverse.

According to approach-avoidance theories, all behaviour can be conceptualised as a response to appetitive (rewarding) or aversive (punishing) stimuli (Elliot & Friedman, 2017). It is generally considered adaptive for humans to approach positive stimuli and avoid negative stimuli. Changes in these approach-avoidance tendencies have been shown to characterise a host of mental illnesses (Loijen et al., 2020). They have also been shown to determine subjective wellbeing (e.g. happiness and life satisfaction, Briki, 2018). Identifying factors that facilitate adaptive and maladaptive approach-avoidance tendencies is thus particularly important for understanding and promoting mental wellbeing.

In exploring associations between WIM-related behaviours and mental wellbeing associated processes, this work sought to understand whether WIM-related behaviours may provide a useful route to mental wellbeing. If confirmed, interventions may be designed around these behaviours and their associated processes. For instance, if reduced social behaviour increased people's approach response towards negative stimuli (a process known to be associated with poor mental wellbeing outcomes) then interventions that train avoidance of negative stimuli in response to reduced social interaction may be tested as a way to bolster against these negative effects. Mapping of such WIM-based processes onto SWB and mental illness respectively can contribute to a better understanding of affective pathways within a broader network of affective change.

Paper 4 " **Reporting on wellbeing-in-the-moment, thoughts and context reduces anxiety"** explores the research question:

- *Does reporting WIM, thoughts and their associated context (activities, location and company), captured during experiential subjective wellbeing measurement, predict improvements in mental wellbeing?*

Also highlighting potential WIM-related pathways to SWB and mental illness intervention, paper 4 explores how two to three weeks of reporting on wellbeing-in-the-moment combined with contextual reports (thoughts, activities and company) as part of a broader experiential subjective wellbeing measurement tool, impacts upon evaluative SWB and mental illness outcomes. The paper consists of three randomised-controlled-trials in which individuals who completed EMA/DRM questionnaires were compared to a control group. The control group in study 1 and 2 consisted of a group that filled out the same EMA/DRM questionnaires without the presumed active ingredient (wellbeing-in-the-moment reports). This "sham" group allowed for isolation of the presumed active ingredient (wellbeing-in-the-moment reports). It also helped to reduce the likelihood that identified effects weren't being driven by positive expectations with respect to the intervention, also known as the Hawthorne effect (McCambridge et al., 2014). The control group in study 3 was a passive control group that completed no EMA/DRM questionnaires at all.

The WIM-related process of interest here was WIM-reporting. It has been theorised that identifying and distinguishing between different feelings helps people better respond to those experiences, since it enables a more specialised and therefore adaptive response (Kashdan, 2015). For example, people who experience their emotions with higher granularity, i.e., using more words to describe both positive and negative emotions, are less likely to use maladaptive coping strategies (such as aggression and self-medicating) and are more likely to use positive emotion regulation strategies that target the specific emotion (Barrett, 2001). Similarly, interventions that train people to improve their emotion differentiation before aversive experiences have been shown to reduce anxiety during those experiences more than cognitive reappraisal and distraction (Kircanski, 2012). It was therefore predicted that contextualising wellbeing reports by complementing them with thoughts and external context reports may positively impact mental wellbeing especially considering extensive evidence that both thoughts and activities are strongly related to wellbeing (Killingsworth, 2010; Smallwood, 2015, Stawarczyk, 2012).

This work provides useful information on measurement in addition to intervention. For example, if WIM reporting has a positive impact on overall mental wellbeing, then it can be usually employed as an isolated intervention. Additionally, if WIM reporting impacts upon on mental wellbeing, this would highlight measurement reactivity as an important artifact of experiential SWB measurement, whereby asking people about how they feel impacts how they feel.

If successful, this work can reveal a relatively light-touch WIM-based intervention that may be added to the suite of tools currently used to improve mental wellbeing. It can also contribute to the mapping of WIM-based processes onto SWB and mental illness respectively contributing to broader theory surrounding their relationships with one another.

1.4. A framework for guiding and synthesising WIM-related mental wellbeing research

All papers in this thesis speak to the association between WIM and overall mental wellbeing. On their own, they represent important contributions to the mental wellbeing literature; but, until they are situated in the context of wider research exploring similar associations, the impact of these studies cannot be fully optimised. By mapping the findings from similar studies onto a common associative framework it will become possible to draw more robust, evidence-based conclusions about which WIM experiences and WIM-based processes and behaviours are associated with which mental wellbeing outcomes. For example, by highlighting the number of studies supporting each association between a given intervention component and mental wellbeing outcome and the strength of evidence supporting each pathway, the different means by which to achieve any given mental wellbeing outcome will become clear. A framework of this kind can provide academics and practitioners with a useful basis from which to better understand the causes and consequences of mental wellbeing.

Crucially, WIM-related mental wellbeing interventions represent a new strand of intervention and research exploring these components separately from other types of intervention components is in its infancy. A framework that allows for the visualisation and synthesisation of this research is a useful first step to developing more focused and robust research in this area with clear theoretical and practical implications.

Therefore, to integrate the findings reported in this thesis and inspire a systematic approach to this new line of research, the discussion section of this thesis concludes by presenting a framework that can be used to map WIM-based intervention components onto mental wellbeing outcomes.

1.5. Summary of key papers

Table 1 below summarises the main papers of the current thesis, their overarching contributions and publication status.

Table 1. Paper contributions and publication status

	Informs measurement	**Informs intervention**	**Publication status**
Paper 1: **The duration of daily activities has no impact on measures of overall wellbeing**	X		Published in Scientific Reports - Nature
Paper 2: **Blurred lines? On the distinction between WIM and mental illness. A two-study replication**	X		In press
Paper 3: **Pandemic-related changes in social interaction are associated with changes in automatic approach-avoidance behaviour**		X	Published in Scientific Reports - Nature
Paper 4: **Reporting on wellbeing-in-the-moment, thoughts and context reduces anxiety**		X	In press

2. Methods

The papers in this thesis employ a variety of methodological approaches which has helped me to develop a broad understanding of the relative merits of each method. Specifically, the papers draw on longitudinal panel datasets conducted within natural (paper 1) or natural experimental (paper 3) settings. They also draw upon online (paper 2) and “in the wild” (paper 4) Randomised Controlled Trials (RCTs). Each paper includes its own methods section where the methods and coding processes used are clearly explained and justified. Please refer to chapter 3, Empirical Work, for these. At the end of each paper, I have also included a final section “Personal reflections” which includes two subsections: “Analysis challenges” and “Key research skills developed”. “Analysis challenges” discusses the difficulties experienced within each paper and how they were overcome. “Key research skills developed” makes clear the research skills that were developed and exhibited during this process.

Since there is some variation in the measures used to approximate WIM and mental wellbeing across the studies the text below briefly outlines which measures were used for each study and why.

2.1. Paper 1 “The duration of daily activities has no impact on measures of overall wellbeing”

Associated research questions:

- *Does the duration of affective moments impact overall (experiential) SWB scores?*
- *Do Ecological Momentary Assessment and Day Reconstruction Method produce comparable SWB reports?*

Key constructs assessed: WIM and Experiential SWB

WIM measurement

In this study WIM was approximated using the specific WIM-based questions housed within the Ecological Momentary Assessment and Day Reconstruction Method. For the EMA, momentary happiness questions “how happy do you feel right now” and purpose “how meaningful does it feel?” were used. The momentary happiness questions for the DRM asked people to recall how they felt in affective moments from the day before, e.g. “how happy did you feel” and “how meaningful did it feel”. These questions were used to approximate WIM since this is standard practice in the EMA and DRM literature and the purpose of this paper was to contribute to that literature. I added the purpose question myself, however, since there had never been any prior explorations of momentary purpose, despite the fact it has been theorised as a key component of SWB.

SWB measurement (experiential)

Non duration weighted SWB was calculated by adding up all responses to “how happy do you feel right now” and “how meaningful does it feel” separately for each questionnaire (EMA and DRM) and then dividing these by the total number of reports. Duration-weighted SWB was calculated using the same approach by weighting the scores by duration of each report. See below for examples. Again, this is the standard practice in the EMA and DRM literature.

Average SWB = $\frac{SWB_1 + SWB_2 + \dots + SWB_Q \; SWB_1 + SWB_2 + \dots + SWB_Q}{Q \qquad\qquad\qquad Q}$, where SWB_i is the reported wellbeing associated to the i-th activity and Q is the number of questionnaires answered;

Average *weighted* SWB =

$$\frac{Dur_1\,SWB_1 + Dur_2\,SWB_2 + \dots + Dur_Q\,SWB_Q \; Dur_1\,SWB_1 + Dur_2\,SWB_2 + \dots + Dur_Q\,SWB_Q}{Dur_1 + Dur_2 + \dots + Dur_Q \qquad\qquad\qquad Dur_1 + Dur_2 + \dots + Dur_Q},$$

where Dur_i is the reported duration of the i-th activity and SWB_i and Q are as above;

2.2. Paper 2 "Blurred lines? On the distinction between WIM and mental illness. A two-study replication"

Associated research question:

- *Does manipulating WIM change anxiety and depression reports?*

Key constructs assessed: WIM and mental illness (anxiety and depression)

WIM measurement – independent variable

In this paper WIM was approximated using the Affective Slider which measures valence and arousal on separate sliding scales ranging from 0-100 (Betella & Verschure, 2016). This measure of WIM was chosen for this study since it quickly captures both arousal and valence dimensions of mood on two slider scales. Since WIM induction was the focus of this study, a quick WIM measure that was unlikely to dissipate or distract from the impact of the WIM induction was necessary. If discrete mood measures were used such as "how happy are you feeling right now", then many emotions would need to be captured for completeness. Measuring WIM in terms of valence and arousal meant that information about the underlying properties of many emotions could be gathered quickly using just two slider scales.

Mental Illness measurement – dependent variable

Mental illness was measured using both trait and state measures of anxiety and depression. Trait anxiety was measured using the trait section of the State Trait Anxiety Index (Gaudry et al., 1975). Trait depression was measured using Becks Depression Inventory (Beck et al., 1961). State anxiety and depression were measured using the State-Trait-Anxiety-Depression-Index (Laux et al., 2018). These measures were selected since they are commonly used and therefore high impact measures of anxiety and depression.

2.3. Paper 3 “Pandemic-related changes in social interaction are associated with changes in automatic approach-avoidance behaviour”

Associated research question:

- *Are changes in WIM-related behaviours (social interaction behaviours) associated with changes in mental wellbeing related outcomes (approach-avoidance tendencies)?*

Key constructs assessed: WIM-related behaviour (social interaction), mental illness associated processes (approach-avoidance), WIM and mental illness (anxiety)

WIM-related behaviour (social interaction) – independent variable

To capture changes in social interaction participants were asked to what extent they had adhered to the prescribed restrictions on social behaviours set out by the UK government during the COVID-19 pandemic. The prescribed restrictions on social behaviours included social distancing, self-isolating, avoiding crowds, avoiding small groups, reduced in-person interaction, reduced overall (digital and in person) interaction) and wearing a mask outdoors.

Mental Illness associated processes (Approach-Avoidance) – dependent variable

Approach-avoidance tendencies have been shown to have a direct association with mental illness. For example, avoidance tendencies are cardinal symptoms of anxiety and are a key mechanism in the maintenance of the condition (Martin, 2022). Depression is also associated with increased avoidance tendencies as well as decreased approach tendencies (Trew, 2011). This paper explored the impact of WIM-related social behaviours on mental wellbeing. Importantly, however, impacts on mental wellbeing may not have been expected to reveal themselves within the timeframe of this study. Previous longitudinal work has found that it takes about 1-2 years for the impact of stressors on mental wellbeing to be realised (Dormann & Zapt, 2022; De Lange et al., 2004). For this reason, mental wellbeing associated behaviours (automatic approach-avoidance tendencies) were selected as the primary mental illness related outcome variable for this study.

WIM measurement

In this study WIM was measured using the Affect Slider which measures valence and arousal on separate sliding scales ranging from 0-100 (Betella & Verschure, 2016). In addition, people were asked to report how happy, sad, and stressed they feel at that moment. Given the respective advantages of both valence/arousal and discrete affective measures we opted for the inclusion of both in this study since brevity was less important for this study. This allowed us to examine any key differences between them within our research context, offering an additional contribution to the affective literature.

Evaluative SWB measurement

The Office of National Statistics 4 wellbeing questions were used as an evaluative SWB outcome measure (ONS-4, 2018). An evaluative SWB measure was chosen since it was beyond the scope and budget of this study to incorporate daily WIM reports that could be used to approximate experiential SWB. ONS-4 was again selected given its wide recognition amongst researchers, policy application, and because it includes both pleasure and purpose dimensions of SWB, both of which are theoretically important (see section 1.1.2).

Mental Illness measurement

Mental illness was measured using the trait section of the State-Trait-Anxiety-Inventory for anxiety (Spielberger, 1983) and the Social Anxiety Scale (Heimberg et al., 1999). Depression was not measured in this study since the two anxiety scales were already quite long and I did not want to risk fatiguing participants. Social anxiety was selected over depression as an outcome measure due to the significance of social anxiety for social behaviours.

Automatic approach-avoidance tendencies were captured using a manikin task (called the Stimulus-Response Compatibility Task, SCRT – Loijen et al., 2020). In the SRCT, participants press a computer key to make a little manikin on the screen move towards (approach) or away from (avoid) the picture. Approach-avoidance tendency scores are computed by taking the difference between the time it takes people to avoid minus approach certain stimuli presented to them. The higher the tendency score, the faster people are to approach instead of avoid a given group of stimuli. For example, the so-called "sad tendency" is people's tendency to approach sad stimuli faster than to avoid them. In this study approach-avoidance tendencies were measured in response to sad/happy faces and social scenes at three different time points over three months.

2.4. Paper 4 "Reporting on wellbeing-in-the-moment, thoughts and context reduces anxiety"

Associated research question:

- *Does reporting WIM, thoughts and their associated context (activities, location and company), captured during experiential subjective wellbeing measurement, predict improvements in mental wellbeing?*

Key constructs assessed: Experiential SWB (WIM-reporting), WIM, Mental Illness (anxiety, depression)

Experiential SWB (WIM-reporting) – independent variable

Experiential SWB was collected as in paper 1 using EMAs and DRMs. These measures were selected since the aim of this paper was to assess whether using popular SWB measures influences mental wellbeing. These questionnaires included daily reports of WIM (happiness and worthwhileness) thoughts, and associated context (activities, location and company) which formed the core ingredient of the intervention.

WIM – dependent variable

To assess impact on measuring experiential SWB on WIM, the momentary happiness and anxiety yesterday items from the Office of National Statistics 4 wellbeing questions were used (ONS, 2018). These questions ask people to report how happy and how anxious they felt the previous day on a scale of 0-10. These questions were selected since they reflect a more momentary assessment of overall wellbeing and mental illness that can be easily contrasted with trait-based measures.

Evaluative SWB – dependent variable

The life satisfaction and worthwhileness items from the Office of National Statistics 4 wellbeing questions were used as single item evaluative SWB outcome measures. An evaluative SWB measure was chosen to determine the impact of WIM reporting on SWB since experiential SWB measures were used as input variables. ONS-4 was selected since it includes both pleasure and purpose dimensions of SWB, both of which are theoretically important (see section 1.1.2). It is also widely used and well-recognised by other researchers and frequently used to inform policy discussions (Dolan et al., 2011). These questions were complemented by two additional questions asking about happiness and anxiety in general.

The World Health Organisation-Five Well-Being Index (Topp et al., 2015) was also included as an additional evaluative SWB measure that may be more subject to change given its focus on a weekly rather than an 'in general' time frame. It consists of 5 items centred around people's experience of positive emotions over the past week, e.g. "I have felt cheerful and in good spirits".

Mental Illness measurement – dependent variable

Mental illness was measured using the Generalised Anxiety Disorder Assessment for anxiety (Spitzer, 2006) and the Patient Health Questionnaire for Depression (Kroenke & Spitzer, 2002). These measures were selected since my study collaborators at Koa Health (online mental health intervention group) had used them in previous intervention research and wanted the results to be comparable. Both are well-established and commonly used scales.

3. Empirical Work

3.1. Paper 1

The duration of daily activities has no impact on measures of overall wellbeing

Amanda Henwood*[1], João Guerreiro[2], Aleksandar Matic[2], Paul Dolan[1]

1. The London School of Economics and Political Science
2. Koa Health

*Corresponding Author

Note: Methods come at the end of this paper due to formatting requirements of the journal to which it was submitted

Abstract

It is widely assumed that the longer we spend in happier activities the happier we will be. In an intensive study of momentary happiness, we show that, in fact, longer time spent in happier activities does not lead to higher levels of reported happiness overall. This finding is replicated with different samples (student and diverse, multi-national panel), measures and methods of analysis. We explore different explanations for this seemingly paradoxical finding, providing fresh insight into the factors that do and do not affect the relationship between how happy we report feeling as a function of how long it lasts. This work calls into question the assumption that doing more of what we like will show up in making us happier, presenting a fundamental challenge to the validity of current tools used to measure happiness.

1. Introduction

It is a substantive fact that feeling happy for longer will make you happier overall. Many behavioural interventions designed to bolster wellbeing focus on increasing the time we spend in pleasurable moments, since it is arguably far easier to control our allocation of time than it is our emotional states. Given the importance of time for happiness, measuring the duration associated with our emotional states has long been considered a staple of happiness measurement. Therefore, the tools we use to measure happiness ought to show a meaningful impact of time on people's happiness reports. But what if current tools used to measure happiness fail to demonstrate this?

Happiness, otherwise known as subjective wellbeing (SWB), refers to how people think and feel about their lives and their everyday experiences (Diener et al., 2002). It is typically measured by capturing two major components: how happy we feel (intensity) and how long we feel it for (duration). Ecological Momentary Assessment (EMA, Stone et al., 1999) and The Day Reconstruction Method (DRM, Kahneman et al., 2004) are two of the most used methods for measuring these components. EMA requires people to report on their happiness at specific moments in time throughout the day (usually selected randomly) alongside the activities that they are engaged in (Stone & Shiffman, 2002). To capture the duration of each happiness episode using EMAs, participants must indicate how long the activity they are currently engaged in has lasted to that point. Instead of relying on "in the moment" reports, DRMs gather wellbeing reports from one point in time relating to a series of episodes that had occurred the previous day (Kahneman et al., 2004). These wellbeing reports follow a diary-like format whereby people report episodes from the day (e.g., working) and how long they last until the entire day is covered.

Changes in SWB have been shown to predict important changes in health-related behaviours (e.g., sleep, Slavish et al., 2018). They are also increasingly being used to assess societal progress on a national scale (Allin, 2007; Dolan & Metcalfe, 2012) and to evaluate policies and programmes (Dolan et al., 2011). Understanding the contribution of various factors, such as duration, to SWB is therefore crucial for advancing both science and society, and a subject of wide appeal to practitioners, clinical and health psychologists, policy makers, economists, and individuals (Krueger & Stone, 2014; Roysamb et al., 2018; Seresinhe et al., 2019)

Evidence has already shown that the proportion of time spent in certain activities can impact overall SWB. For example, the proportion of time spent in happy relative to less happy activities has been shown to be an important determinant of overall SWB (Diener et al., 2009). There is, however, currently a dearth of empirical research exploring whether current SWB measures are picking up on one of the most fundamental assumptions: namely, that more time spent in happy activities is better for us than spending less time in happy activities. SWB measures must meet this assumption if they are to reliably inform behavioural interventions targeting happiness.

The present empirical research

To assess the impact of duration on reported happiness, we gathered daily EMA and DRM happiness reports using a mobile app over a 2–3-week period in two large, and diverse samples: one mixed (people of different ages, gender and nationality) and one student sample. Since SWB is determined by the purpose (or worthwhileness) of an activity as well as by its pleasure (or happiness), participants were asked to report both the worthwhileness and happiness in both EMA and DRM (Dolan, 2014). The required sample size to detect a medium effect size (0.50) with a high power criterion (90%) and standard significance level (α = 0.05) when comparing duration with non-duration weighted SWB is 43 (Faul et al., 2007). However, since we planned on exploring other variables as moderators, and to aid replicability of our results with an extended set of registered hypotheses[1], we aimed to recruit as many participants as possible. Sensitivity power analysis with the final number of participants (217 for the student

[1] We preregistered our hypothesis through the Open Science Framework (OSF; https://osf.io/yt745)

sample and 195 for the mixed sample) included in our analyses for the core hypothesis (ref section to the Appendix), suggested that our study had the power of 90% (α = 0.05) to detect Cohen's d of 0.22 and 0.20 respectively, which is close to the small effect of 0.20.

The two studies were conducted in a natural setting as opposed to a lab for two reasons. First, an impact of WIM reporting was not expected to take place immediately, and as such, SWB measures had be completed over time to assess their impact on overall mental wellbeing. It is not feasible to conduct such a study in the lab. Second, experiential SWB measures are typically completed in a natural setting as people go about their daily lives. As such, the research validity would have been significantly reduced by assessing their impact in a lab environment.

Analysis plan

Simple t-tests comparing the measure (EMA vs DRM) and then the method (duration vs non-duration weighted scores) for each happiness and purpose question was all that was necessary to answer the main research questions for this paper. However, for completeness we also chose to explore whether the average similarities we observed across measures and methods persisted at the individual level by using scatterplots of individual responses. For additional robustness, we also checked to see whether the similarities observed persisted using different methods of experiential SWB calculation (daily averages of WIM scores vs total averages of WIM scores).

For significant effects, we considered values above a 1% difference in SWB (0.1) to be a meaningful difference. Other papers have reported a difference of around 5% for major life events such as unemployment and disability, and so 1% makes it more likely that differences will be found (Dolan & Peasgood, 2008).

Ethics statement

All studies were conducted in line with GDPR and the guidelines of the American Psychological Association. All participants gave informed consent on joining the study and the experimental protocol was approved by the London School of Economics Ethics Committee. The study did not involve deception and hypotheses, methods, and analysis for the second of the two studies were pre-registered. We applied many different robustness checks to ensure robustness of our findings, which can be found in the Appendix.

2. Results

In analysing these data, we first report descriptive statistics of SWB intensity scores for each sample, method, and measure (see *Table 1*). These are presented in the table below.

Table 1. Descriptive statistics of SWB intensity scores for each sample, method, and measure, with confidence interval at 95% level.

Mean intensity	Student sample		Mixed sample	
	EMA	DRM	EMA	DRM
Happiness (CI@95%)	7.17 ±1.83 [7.14, 7.20]	7.17 ±1.76 [7.14, 7.21]	6.44 ±1.91 [6.40, 6.48]	6.53 ±1.98 [6.49, 6.58]
Worthwhileness (CI@95%)	7.34 ±1.90 [7.30, 7.37]	7.42 ±1.77 [7.39, 7.46]	6.39 ±2.19 [6.34, 6.43]	6.57 ±2.16 [6.52, 6.63]

We note that mean happiness intensity scores were slightly lower for EMA than for DRM in the student sample. The average reported happiness intensity for EMA and DRM in the student sample was (*EMA:* M=6.44, SD=1.91, CI(95%)=[6.40, 6.48]; *DRM:* M=6.53, SD=1.98, CI(95%)=[6.49, 6.58]). This difference was not present in the mixed sample. The average reported happiness intensity for EMA and DRM in the mixed sample was (*EMA:* M=7.17, SD=1.83, CI(95%)=[7.14, 7.20]; *DRM:* M=7.17, SD=1.76, CI(95%)=[7.14, 7.21]). Mean worthwhileness intensity scores were also slightly lower for EMA than for DRM in both samples. The average reported worthwhileness intensity for EMA and DRM in the student sample was (*EMA:* Mean=6.39, SD=2.19, CI(95%)=[6.34, 6.43]; *DRM:* Mean=6.57, SD=2.16, CI(95%)=[6.52, 6.63]). The average reported worthwhileness intensity for EMA and DRM in the mixed sample was (*EMA:* Mean=7.34, SD=1.90, CI(95%)=[7.30, 7.37]; *DRM:* Mean=7.42, SD=1.77, CI(95%)=[7.39, 7.46]). Overall, the mixed sample reported higher happiness and worthwhileness intensity than the student sample.

In terms of duration, the average reports for EMA and DRM in the student sample were (*EMA:* Mean=154 min, Median=120 min, SD=118 min, CI(95%)=[152, 156]; *DRM:* Mean=163 min, Median=120 min, SD=135 min, CI(95%)=[160, 166]). The average reported duration for EMA and DRM in the mixed sample was (*EMA:* Mean=158 min, Median=100 min, SD=143 min, CI(95%)=[156, 160]; *DRM:* Mean=168 min, Median=120 min, SD=151 min, CI(95%)=[165, 170]). Note that we doubled the EMA duration times to account for the fact that people are on average interrupted mid episode and enable comparison between measures. Taking these doubled EMA reports into account, DRM duration reports are about 10 mins longer than EMA duration reports in both samples (statistically significant with p-value < 0.001). Although significant, this is a small difference and likely due to differences in the recall style of each reporting method: EMAs rely on in the moment recollection whereas DRMs rely on recollection of the previous day. Further exploration of this difference is beyond the scope of the current paper.

The results speaking to the main research question - *how does duration-weighting contribute to overall SWB?* - are presented in *Figures 1 and 2*. For brevity, we report only results for happiness in the main results section below (*see Tables in Appendix* for worthwhileness). The graph in *Figure 1* shows that overall SWB is very similar whether or not duration is accounted for. This holds for both measurement methods and both samples. This similarity also holds at an individual level, as evidenced by *Figure 2.*

The results show that overall SWB does not change by more than 1% whether or not duration is weighted. This result holds even when different methods of SWB are employed, such as when using daily averages of SWB reports vs. total averages of SWB over the course of the 2-3 week study (*see Statistical Analysis in Appendix* for specific calculations).

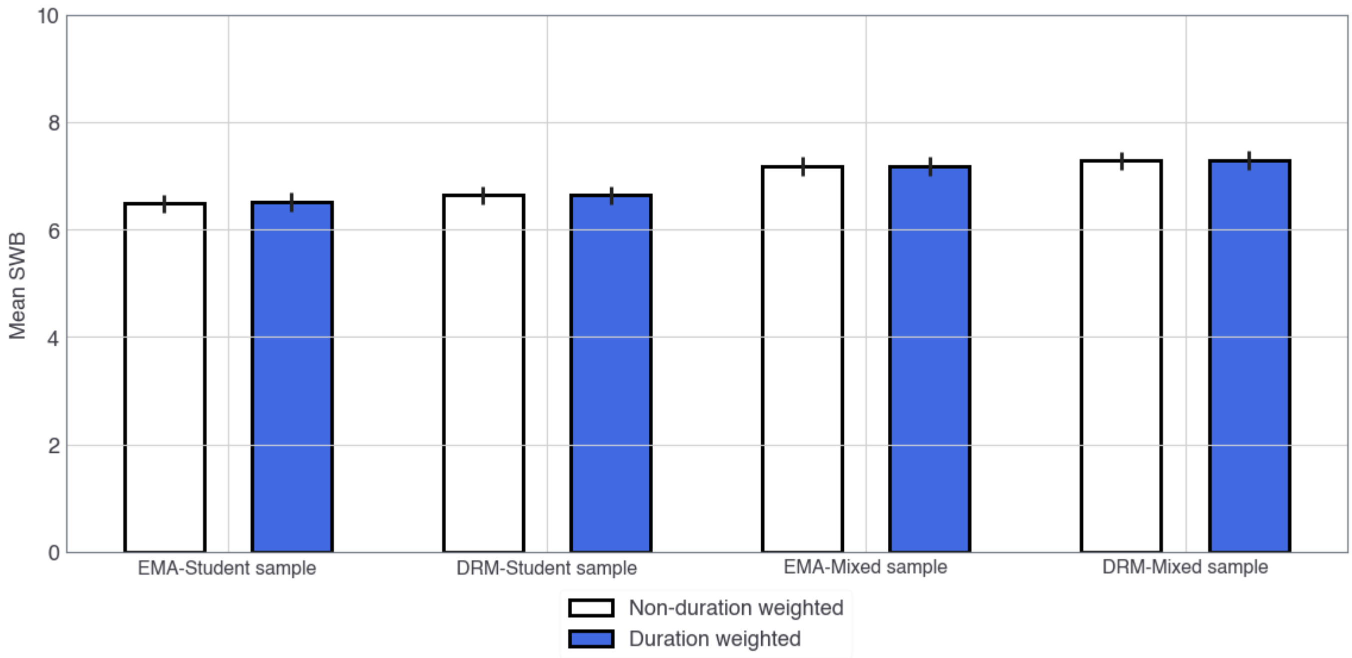


Figure. 1. *Mean SWB (happiness intensity x duration) scores with and without duration weights, for each questionnaire type and sample*

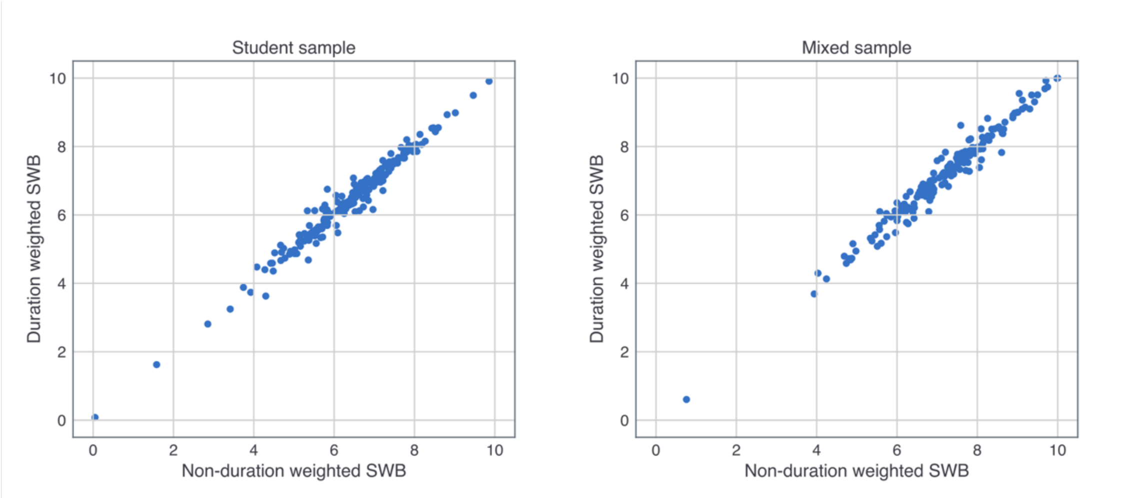


Figure 2. *Scatter plots showing total reported SWB (happiness intensity x duration) scores per individual, with and without duration weights, reported in the EMA questionnaire. The figure on the left represents the student sample and the figure on the right represents the mixed sample.*

Table 2. Mean of pairwise differences between average reported SWB (happiness intensity x duration) with duration weights and average happiness without duration weights, with confidence interval at 95% level. Significant results mean that mean difference is less than 1%. (*** if p-val < 0.001, ** if p-val < 0.01). (see Appendix, Table S.13 for worthwhileness)

Mean difference	**Student sample**				**Mixed sample**			
	EMA		**DRM**		**EMA**		**DRM**	
	Total averages	**Daily averages**	**Total averages**	**Daily averages**	**Total averages**	**Daily averages**	**Total averages**	**Daily averages**
Happiness (CI@95%)	0.035*** [0.005, 0.064]	0.028*** [0.014, 0.042]	-0.007*** [-0.038, 0.023]	0.004*** [-0.012, 0.020]	-0.008*** [-0.041, 0.025]	0.005*** [-0.009, 0.018]	-0.021*** [-0.050, 0.008]	0.005*** [-0.014, 0.023]

We conducted post-hoc analyses to explore several factors that might explain why weighting by duration might not affect overall SWB. First, we explored variation in the two core variables: intensity and duration. We found that for individuals with daily activity patterns that result in higher variance in duration (*see Appendix, Table S.1*), and/or in high variance in intensity reports (*see Appendix, Table S.2*), the differences between duration-weighted and unweighted scores became greater, such that all the t-tests with respect to the total averages yielded results that were not statistically significant and thus greater than 1%.

Second, we explored the relationship between the two core variables: intensity and duration. For each person, we sub-sampled reports by selecting those that had a relatively higher correlation between intensity and duration. These reports were sampled by randomly sampling 100 sets containing half of each person's reports and then picking the set with the highest correlation between intensity and duration. Interestingly, we found that duration did impact overall SWB for reports with these higher correlations: the higher the percentile of correlation (whether it was positive or negative) the greater the difference between weighted and non-weighted SWB (*see Appendix Table S.3*). Moreover, the correlation was found to be the most important driver in explaining the difference between duration weighted and unweighted SWB, in contrast with the standard deviation of the SWB and duration variables (*see Appendix, Tables S.4-S.5* for details of the regression analysis). These three variables explain between 45% and 81% of the variance observed in the difference variable. We also note that it was not possible to increase the strength of the correlation between intensity and duration by focusing on either overall SWB reports, or SWB reports associated with specific activities, that were higher (happier) or lower (more miserable) (*see Appendix, Tables S.6, S.7-S.10 + S.11, S.12*).

For robustness we also checked whether the difference between duration-weighted and unweighted SWB changed as a function of different demographic and interpersonal characteristics including high and low absolute happiness levels per person, personality, age, education, employment, gender, income. None of these variables explained the lack of relationship between duration and overall SWB. We also found no strong evidence for the possibility that happy activities make people happy up until a certain point as would be predicted by the law of diminishing marginal returns. For example, 30 minute episodes had higher mean intensity than 1 hour episodes but 1 hour 30 minute episodes had an even higher intensity. If duration were dissipating the impact of intensity would expect activities with longer durations be less intense in general.

3. Discussion

In an exploratory study, and in a follow up pre-registered experiment using diverse, multi-national samples, we found that weighting SWB reports by the duration of daily activities does not change overall SWB. This research challenges the assumption that the duration of emotional episodes makes a meaningful contribution to the calculation of SWB, calling into question the validity of current measurement tools. The result is robust to differences in SWB measurement, SWB calculation method (average of daily or total reports), high/low happiness experiences, as well as different demographic and interpersonal characteristics.

Our data highlight several possible explanations for this result. It appears that the low correlation between happiness intensity and duration reports is predominantly driving the result; reports with higher correlations (positive or negative) between intensity and duration yield larger differences in duration weighted and non-weighted SWB scores. More variation in scores typically generates stronger correlations. Indeed, in our data we found higher variance in intensity and/or duration reports to be associated with larger differences between duration and non-duration weighted happiness. Duration is (understandably) not expected to influence overall SWB if activities are always lasting roughly the same amount of time on average and/or if people report roughly the same happiness intensity on average.

These results call into question the reliability of existing happiness measures. Whilst it is possible that the low variation in duration and intensity (and therefore the lack of relationship between the overall SWB and duration) is a true reflection of how these entities vary in real life, this result is perhaps more likely to be a product of mismeasurement. In capturing duration associated with activities rather than emotional experience, it remains a possibility that current measures may fail to capture important variance in duration that is related to SWB. For example, although you may report feeling 4/10 happiness whilst commuting, that feeling might have been influenced by how you felt just before you started commuting. This "emotional lag" across activities means that the duration of this emotional episode may not be captured completely by focusing on commuting alone. Indeed, in our exploratory analysis, we show that happiness from the last emotional episode is a stronger predictor of current SWB reports than duration from the previous emotional episode. This lends support to the idea that the beginning and end of emotional experiences are not always clearly signalled by the beginning and end of any given activity – carryover effects are present (*see Appendix, Table S.14*). Further manipulation studies where these two types of happiness measures (activity and emotion based) are directly compared will be necessary to affirm this possibility.

Moreover, associating duration with activities as a proxy for the duration of our emotional experiences may complicate the relationship between SWB intensity and its duration. Spending more time in an *activity* we like may start to yield less happiness after a while, whilst this is less likely to be the case for emotional experiences. A longer time spent feeling happy is unlikely to make us feel any less happy beyond a certain time since we are already capturing a direct measure of emotional experience. Importantly, previous research has identified a strong relationship between the duration of emotional states and their intensity when emotion is being measured directly. In a study where participants were asked to provide daily reports on their experiences of anger, joy, or fear, and rate their intensity, higher emotional intensity was found to be significantly associated with longer durations for all three emotions (Verduyn et al., 2009). Thus, a focus on activity duration instead of happiness duration may be obscuring the relationship between happiness duration and intensity.

Interestingly, however, it is also possible that the reason for the observed similarity between duration and non-duration weighted SWB is that people implicitly take their SWB into account when reporting the happiness levels associated with their current activities. If this is the case then it may not be necessary to capture duration since the SWB report will already have this detail captured implicitly. Future work testing the extent to which SWB reports associated with the same activity are impacted by how long that activity lasts.

Conclusions and limitations

This research provides substance to the concerns of those already questioning the validity of self-reported happiness (Diener et al., 2009) and should concern academics and practitioners that use SWB tools to evaluate impact. Providing that these findings are replicated in further studies, including those that directly compare activity and emotion based SWB reporting, new happiness measures may need to consider additional sources of measurement to complement, or perhaps even replace, self-report. For instance, identifying the exact start and end point of our emotional experiences will be a more cognitively demanding task than identifying the start and end point of activities. Therefore, affective computing approaches that generate an automatic mapping of additional variables predictive of certain emotional states for participants may yield more promising results in terms of capturing the full trajectory of emotional experience (Voukelatou et al., 2020). They will also be better able to detect other factors that may increase the variability of emotion duration such as emotional triggers (Verduyn et al., 2009). For example, studies using smartwatches that detect heart beats and light exposure have found that happiness has an important association with these parameters (Gloor et al., 2018). Against this background, rapidly evolving new technologies that enable passive monitoring of these additional variables represent a promising avenue for more effective SWB monitoring tools in the future (Poria et al., 2017; Tau & Tan 2009; Picard, 2005).

In this study we were unable to discern the causal impact of differing reported happiness levels on the duration of activities and vice versa. This would require a manipulation study, where ecological validity would be reduced and exploration of the relationship between intensity and duration with respect to the most popular measurement tools would not be possible. Considering these findings this would be useful complementary research. Importantly, however, by focusing on EMA and DRM reports over the same period by the same people over time, our study design allowed for the isolation of differences in SWB at the episode level (how the same person reports emotion intensity and duration across measures in the same episode) and at the daily level (how the same person reports emotion intensity across measures on the same day). These interpersonal comparisons are critical for generating more robust and reliable approximations of SWB (Kristoffersen, 2010). Despite the diversity of samples used in this study, we recommend that replications of this study with both similar and additional sub-samples (such as those with mental health problems) could be conducted to further improve generalizability of these results. Finally, future work should explore the impact of self-reported vs objective durations within in this context since self-reporting of duration may be impacted by individuals SWB at the time.

Going beyond previous work focused on how the frequency of positive experiences contribute to overall SWB, we show that duration does not meaningfully contribute to the calculation of SWB using existing measures. Our results do not rule out the fact that longer time spent in happy experiences is good for overall SWB. However, they do cast doubt on the ability of existing wellbeing measures to show that this is true. The reasons for this must take on an important new line of scientific research and perhaps new happiness measurements.

4. Methods

Participants

Phase 1 of the data collection was conducted in Spain, Colombia, Chile, Peru and UK (mixed sample). This sample was recruited between the dates 25/05/2018 and 24/09/2018 through a recruitment agency who attempted to diversify the sample with respect to age, gender, education, and socio-economic status. Phase 2 was conducted in the UK (student sample) between the dates of 29/12/2018 and 03/03/2019. The student sample were from the London School of Economics University in England and were recruited via authorised university communication channels including social media, university newsletters and email. The demographics for each sample can be found in *Table 3* below. All participants signed a consent and privacy policy form on entering the study. Participants were only allowed to complete the study if they were: over 18, able to use a smartphone, and had no current mental health diagnosis (to control for the potential impact of mental health related medications). Only Android and iPhones were allowed in the study. However, since together these brands account for 99% of the global market share and this figure will be higher in the countries specified, we consider this to be a good representation of the population.

In this analysis, we started from a sample of 582 in the mixed group and 653 in the student group who had completed mood reports since this was the focus of our paper. From this sample, we excluded SWB reports where duration lasted longer than 12 hours, as well as EMA reports where there appeared to be duplicates (two episodes reported within less than 15 minutes of each other). We included in the analysis the participants who submitted more than 20 EMA reports and at least 5 DRM reports in total, which implies that they were active for at least 5 days of the study. These criteria resulted in the exclusion of 387 participants from the mixed sample and 436 participants from the student sample. The final sample numbers are listed in the *Table 3* below.

Table 3. Descriptive statistics of study samples

	Student sample	**Mixed sample**
Number of participants	217	195
Mean age	22.9 ± 3.7	31.5 ± 6.1
Mean income	£665 ± 555	€1964 ± 1073
Gender split	62% female, 38%	58% female, 42%

	male	male
Employment	20% employed, 80% not employed	51% employed, 49% not employed
Life satisfaction	6.82 ± 1.66	7.16 ± 1.55

Procedure

As part of a larger study exploring the determinants of SWB participants were instructed to download a custom designed mobile app via a specially curated study webpage. Once downloaded, participants underwent a series of initial onboarding questions via the app including questions relating to overall life satisfaction, worthwhileness, daily happiness and anxiety, and overall happiness and anxiety. These questions were followed by demographics and trait-based questionnaires including personality (50 item Big Five personality assessment questionnaire available on the IPIP website[2]). Participants then entered a SWB monitoring period of 2-3 weeks within the app during which they completed five daily EMA and once daily DRM reports, alongside several other SWB measures. The app notified them when responses were necessary. Given the high intensity of SWB reports required, we considered this time frame as being sufficiently long enough to show change but not too long that too many of the sample would be lost. At the end of the study respondents were asked to complete the same initial questions relating to overall life satisfaction, worthwhileness, daily happiness and anxiety, and overall happiness and anxiety, that they received in the onboarding section. An overview of the procedure for studies 1 and 2 is depicted in Figure 3 below.

Participants were able to report bug-related concerns via the app anonymously. These were responded to in app and anonymously by the assigned Alpha employees. Respondents with over 70% completion rate were reimbursed with £20 Amazon

[2] Available at http://ipip.ori.org/New IPIP-50-item-scale.htm. Accessed on December 2011.

vouchers (student sample) and £40 (mixed sample). We decided that these incentives would be large enough to recruit a big enough *n* in the respective samples, but not sufficiently high enough to change people's wellbeing. The higher price in the mixed sample was due to rates set by the recruitment agency used and reflective of a mostly working non-student sample.

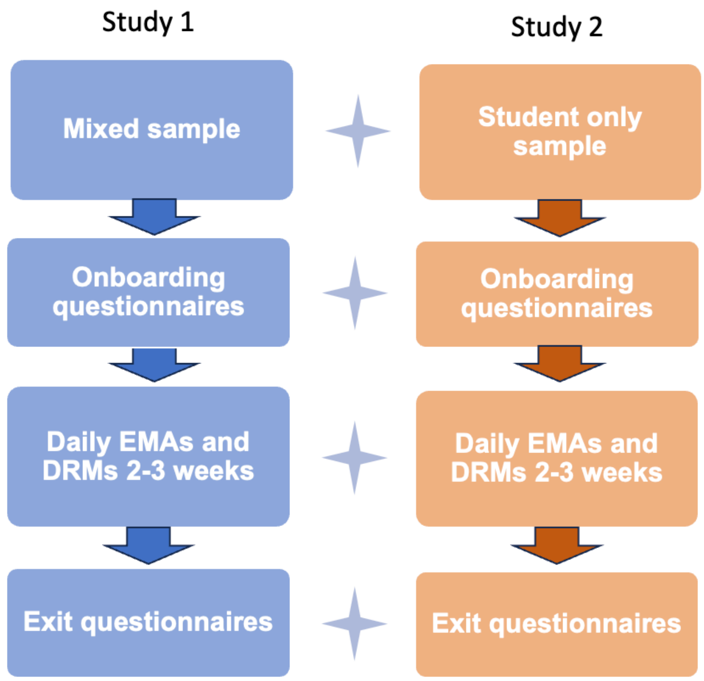


Figure 3. *A schematic diagram depicting an overview of the procedures for studies 1 and 2. The key difference here between the two studies is the sample: study 1 is a mixed ethnicity sample and study 2 is a student only sample from the UK.*

Measures

SWB measured by Ecological Momentary Assessment. EMA reports consisted of responses to various prompts issued at five random intervals throughout the day. First, participants had to select an activity (e.g. "working") from a list of common activities in response to the prompt: "During the past hour, I was". Activity lists differed depending on whether the sample was student or mixed. The student sample received common activities (e.g. "eating") in addition to activities that were tailored to university life (e.g.

“studying”). The mixed sample only received common activities. Next participants had to indicate the duration of the activity so far using a drop-down tab which showed time periods that went up in 10-minute increments, ranging from 10 minutes to 4 hours and 10 minutes, in response to the prompt: “How long have you been doing this?”. Then participants had to indicate who they were with, what they were thinking about, and where they were, from a list of common suggestions (e.g. “Kids”, “Events from my past”, “At my parents’ house”) in response to prompts: “I was with”, “I was thinking about”, “Where are you?”. Finally, participants had to answer how they felt on a scale of 0-10 in response to prompts: “How happy did you feel?” and “How worthwhile did this feel?”. The timing of EMAs was randomised.

SWB measured by Day Reconstruction Method. Every morning, participants were asked to provide an overview of the previous day partitioned in episodes. We used the text from the DRM instructions provided in *Kahneman et al*, 2004: “Think of your day as a continuous series of scenes or episodes in a film. Give each episode a brief name that will help you remember it (for example, `commuting to work', or `at lunch with B'...). Write down the approximate times at which each episode began and ended.”

Participants then had to indicate what they were doing (e.g. “working”) in response to the prompt “I was doing” and select a “start time” and “end time” for the episode. Activity lists differed depending on whether the sample was student or mixed, as described above in EMA reports. Also, like EMA -reports, for each episode participants had to indicate who they were with and what they were thinking about in response to prompts: “I was with” and “I was thinking about”. Finally, as per EMA reports, participants had to answer how they felt on a scale of 0-10 in response to prompts: “How happy did you feel?” and “How worthwhile did this feel?” Participants could not complete the DRM without having covered 12 hours of emotional episodes.

SWB calculations. For robustness, we used four different formulas for calculating SWB: 1) average SWB scores aggregated over the full length of the 2-3 week studies

(Average SWB), 2) total average SWB scores aggregated over the full length of the 2-3 week studies weighted by duration (Average weighted SWB), 3) average of daily SWB scores (Daily average SWB), 4) average of daily SWB scores weighted by duration (Daily average SWB weighted). See below for details.

Average SWB = $\frac{SWB_1 + SWB_2 + \ldots + SWB_Q SWB_1 + SWB_2 + \ldots + SWB_Q}{Q \qquad Q}$, where SWB_i is the reported wellbeing associated to the i-th activity and Q is the number of questionnaires answered;

Average *weighted* SWB =
$\frac{Dur_1 SWB_1 + Dur_2 SWB_2 + \ldots + Dur_Q SWB_Q Dur_1 SWB_1 + Dur_2 SWB_2 + \ldots + Dur_Q SWB_Q}{Dur_1 + Dur_2 + \ldots + Dur_Q \qquad Dur_1 + Dur_2 + \ldots + Dur_Q}$,

where Dur_i is the reported duration of the i-th activity and SWB_i and Q are as above;

Personal reflections on paper 1

These are personal research reflections and do not form part of the main paper. The purpose of this section is to highlight analysis challenges and key research skills developed.

Analysis challenges

The most significant challenge encountered when analysing these data was in trying to understand the reason that duration did not contribute to overall SWB scores. To explore possible reasons several post-hoc tests were conducted such as seeing whether SWB scores were affected by duration in cases with particularly high or low duration scores and in people with particularly high or low evaluative SWB to begin with. Demographic factors, such as age or gender, or personal factors, such as personality, were explored as potential driving factors behind these results.

Deciding how many WIM reports were necessary to approximate SWB was also a challenge. It was decided that if there were enough reports to account for roughly 5 days' worth of affective experiences as a minimum this would be sufficient for assessing impact. This decision was based on the fact that 5 days was already an advancement on previous research comparing these measures and still feasible as a minimum threshold considering the size of our dataset.

Key research skills developed

This paper relied on two large-scale experimental studies that I played a lead role in designing and collecting the data for. The studies captured daily WIM reports using a custom designed mobile phone app. I led the development of this application, coordinating insights from computer scientists and app developers to produce the final product. To promote the app and get as many people as possible to use it I developed a recruitment plan that utilised LSE's social media channels and student networks.

I was solely responsible for the write up of the paper, the literature review, and the paper publication process. I also played a significant role in shaping the study conception, analysis plan, analysis decisions and analysis interpretation.

3.2. Paper 2

Blurred lines? On the distinction between WIM and mental illness. A two-study replication

Amanda Henwood*[1,]

1.The London School of Economics and Political Science
*Corresponding Author

Abstract

Despite the well-established, strong positive association between wellbeing-in-the-moment (WIM) and mental illness, little is known about the extent to which these constructs overlap. Without this knowledge the foundation upon which we base our understanding of this association and the practical implications that follow from this remain weak. This paper reports the results of two pre-registered online RCTs assessing the impact of changes in WIM on trait anxiety and depression (study 1) and both trait and state anxiety and depression (study 2) questionnaire responses. Prior to answering anxiety and depression questionnaires, participants underwent WIM inductions that were either sad (negative de-aroused), anxious (negative aroused), happy (positive aroused), relaxed (positive de-aroused) or neutral. In both studies, scores on trait and state anxiety scales remained remarkably stable across WIM conditions. Some overlap was identified, however, with respect to negative affectivity and trait/state depression. The implications of these findings for anxiety and depression measurement, as well as theoretical understandings of their association, are discussed.

1. Introduction

Mental illness is one of the biggest determinants of human suffering with profound additional health and economic consequences (Firth et al., 2019; McDaid et al., 2019). WIM is an important, and largely related, construct to mental illness (Ando, 2002). There is an ever-increasing body of research highlighting the importance of WIM in determining mental illness (Forrest et al., 2021; Geschwind et al., 2011; Shin et al., 2021). Recent work has even suggested that WIM and mental illness measures can be used as substitutes for one another (Targum et al., 2021). Yet despite their strong relationship, little is known about the extent of overlap between these constructs. Assessing overlap can help to determine whether the strong relationship between WIM and mental illness exists due to a meaningful interaction between the two constructs or because both measures pick up on common aspects of WIM. Gathering such insights can help affirm the validity of current measures used to capture these constructs, advance theoretical understanding of why they relate to each other, and clarify whether similar mental illness constructs such as anxiety and depression share similar WIM-based components.

According to the Tripartite Model of anxiety and depression, WIM and mental illness sit on a continuum whereby anxious and sad moods lie at one end and severe anxiety and depression lie at the other (Chentsova-Dutton et al., 2015). This type of conceptualisation is consistent with disease models of anxiety and depression. If this is the case, then any negative affect experienced at the time of reporting anxiety and depression may reveal itself in those scores, leading to an inflation of mental illness reports. This is a problem for measurement since to achieve a valid measure of mental illness that is consistent across and within people over time it is important that this measure is not biased by transient features such as WIM (Watson et al., 1995). Conflation (or otherwise) of WIM and mental illness has important implications for researchers seeking to understand the relationship between these two constructs (Shaffer et al., 2016). This is because any degree of overlap between two measures can distort estimates of the relationship between the two underlying constructs (DeVellis

& Thorpe, 2021). That is, a relationship might be detected, not because the two constructs are related, but because some of the variance of each of these measures relates to the same construct to begin with (e.g. both measures capture aspects of negative WIM).

An impact of changes in WIM on mental illness reports could result in inaccurate mental illness comparisons across people and administration of mental health treatment based on most in need in the moment, as opposed to most in need overall. From an intervention perspective, it may be difficult to decipher whether changes in mental illness observed in response to an intervention should be attributed to the intervention itself, or to changes in WIM that the intervention might have generated. An understanding of whether WIM and mental illness can be meaningfully separated from one another is also key to understanding how these two constructs might contribute to arguably broader constructs, such as mental wellbeing (Van Agteren et al., 2021). For example, if they can be separated then it becomes feasible to consider WIM and mental illness as independent pathways to wellbeing, and thus ensure that wellbeing interventions contain components that target both.

Critiques of the Tripartite Model have argued that conflation of WIM and mental illness is harmful since anxiousness and sadness are moods, but anxiety and depression are disorders. Confusion of these constructs can impair the conceptualisation, diagnosis and treatment of anxiety and depression (Chentsova-Dutton et al., 2015). It may also result in mental illnesses not being taken seriously as disorders. However, before a stance can be taken, it is important to know to what extent, and under what conditions, overlap between WIM and mental illness exists.

Importantly, if multiple mental illness measures (e.g., anxiety and depression) are affected by negative feelings at the time of completion, this could also lead to empirical conflation of those measures. That is, an inflated relationship between anxiety and depression may be observed due to the fact they are both picking up on a common underlying factor: negative WIM. Anxiety and depression are commonly conceptualised as sharing a large proportion of negative affectivity (Clark & Watson, 1991). According to these conceptualisations, it makes sense for anxiety and depression measures to be influenced by negative WIM in similar ways. Though, as explained, this may lead to measurement complications. Testing this assumption, as well as establishing the relative extent of negative affectivity in each condition (anxiety/depression) and whether this negative affectivity is distinct from short-term experiences of mood, remains a critical consideration in affective research.

To address this outstanding uncertainty, the present work tests whether WIM manipulation alters people's responses to two types of mental illness questionnaires (anxiety and depression) in two online studies. The work focuses on the psychological disorders, anxiety and depression, for three reasons. First, they are the most common mental illnesses and therefore have a high prevalence in the population (Chalder et al., 2012). Second, they are disorders of mood, which means they may be more likely than other mental illness questionnaires to share variance with WIM. Third, they are highly correlated with one another, and many researchers have questioned their separability as constructs (Dobson, 1985).

"WIM, anxiety and depression" section below provides definitions of WIM, anxiety and depression and provides an overview of existing literature detailing their relationship with one another. "Associative responding" and "empirical overlap" sections present two key theory-based explanations for how mental illness measures may be influenced by WIM. Section 2 outlines the details of the present research. Section 3 and 4 outline the two studies conducted to explore whether WIM impacts mental illness reports, including methods, results, and individual discussions. The paper finishes with Section 5, which is a general discussion combining the insights from both studies.

WIM, anxiety and depression

Affect is a general term used to describe emotional phenomena, which encompasses both emotion and mood (Frijda, 1993). Emotions are typically brief, intense, and directed towards a stimulus whereas moods tend to have more uniform and reliable consequences and are relatively more enduring (Forgas, 2013). All affective phenomena can be measured across two key dimensions: valence (how pleasant someone feels) and arousal (how activated someone feels) (Barrett & Russell, 1999). Due to their relatively enduring nature, this study assumes that moods will be most likely to share common variance with mental illness and thus focuses on the relationship between mood and mental illness. Mood is used here as a means of capturing WIM (defined earlier in PhD introduction).

Anxiety is an affective, physiological, cognitive, and behavioural state (Gaudry et al., 1975). It stems from perceived threats to future happiness, self-esteem, and sense-making (Dobson, 1985). Whilst fear relates to the awareness of dangers, anxiety is the accompanying affective state and physiological reaction that occurs in response to fear. State anxiety – anxiety experienced in the moment - is a negative affective state characterised by high arousal (activation) levels. Trait anxiety, on the other hand, refers to the tendency to become anxious and perceive things as threatening, which may or may not be accompanied by anxiety in the moment (Baker & Guttfreund, 1993). As such, trait anxiety questionnaires tend to ask questions about the frequency or extent to which people generally experience certain symptoms of anxiety, e.g. "I feel nervous and restless".

The separability of state and trait anxiety has been demonstrated empirically in studies showing that anxiety manipulations produce changes in state but not trait anxiety (Gaudry et al., 1975). However, this work did not assess the impact of sad mood manipulations on state/trait anxiety. It also induced anxiety by communicating success or failure on an anagram task and thus the focus was on externally, as opposed to internally, generated anxiety. One study that did test the impact of internally generated WIM on state anxiety has reported a significant impact. This study induced negative/positive WIM (sad/happy mood) using an autobiographical recall task which instructed participants to either think about events that made them feel the most lonely, sad, rejected, or hurt or think about the two happiest events of their lives (Baker, 1993). It found that increases in sad mood produced increases in state anxiety and increases in happy mood produced decreases in state anxiety. However, it did not explore the impact of internally generated anxiety induction on state anxiety. Thus, research assessing the impact of WIM manipulations across the spectrum (high-low valence and high-low arousal) on trait and state anxiety can help to address these voids in present knowledge.

Like anxiety, depression is also considered to have affective, physiological, cognitive, and behavioural components. It is associated with threats to self-esteem, happiness, and ability to cope – though, unlike anxiety, these threats are considered imminent, certain, or as having already happened (Dobson, 1985). Depression questionnaires ask people to indicate whether, or to what extent, they are experiencing certain enduring symptoms currently or during the previous week, e.g. “I am sad all of the time and I can’t snap out of it”. Emotion theories have used differences in activation as key defining factors that allow for a distinction between anxiety and depression. For example, Russell’s circumplex model of emotion positions 28 emotion adjectives in a circular array across two key dimensions (pleasure-displeasure and low-high arousal) (Russell, 1980). In this model, depression is placed on the low arousal, high displeasure dimension whilst anxiety is placed on the high arousal high displeasure dimension.

Other emotion theories have associated anxiety and depression with key emotions. For example, Differential Emotions Theory explains that fear is a key emotion in anxiety with which other emotions interact and distress/anguish (which predominates in grief) is a key emotion of depression with which other emotions interact (Izard, 1977). Anxiousness has been said to include curiosity, alertness, and distress (also defined as sadness), whilst depression also includes elements of fear, hostility and distress. Both theories suggest significant crossover, but also diversion, in emotional experience between depression and anxiety and notably do not clearly distinguish them from momentary affective states. This leaves open the question of whether changes in WIM are in fact distinct from changes in anxiety and depression.

Prior work has documented an association between changes in WIM and trait depression (Peeters, Berkhof, Delespaul, Rottenberg, & Nicolson, 2006). Other work has found that mood becomes more variable in the presence of depressive symptoms (Bowen et al., 2013). However, these studies are associative, and thus it has not been possible to determine whether such associations arise due to shared variance between WIM and mental illness or whether they simply coexist. Manipulating WIM experimentally is the most rigorous way to test the causal impact of WIM on any given variable (Siedlecka & Denson, 2019). Research exploring the impact of WIM manipulations on trait anxiety and depression reports is therefore key to addressing these gaps in knowledge.

Both state anxiety and depression are more momentary, and context based than trait anxiety/depression (Yaden & Haybron, 2022). However, their separation from WIM as constructs suggests that they are still considered to be distinct from general negative affect. As part of the aforementioned study that used an autobiographical recall task to assess the impact of happy/sad mood on state anxiety, the same study explored the impact on state depression, finding a significant impact in a negative direction for happy and a positive direction for sad mood (Baker, 1993). Notably, however, this research did not explore the impact of anxious or relaxed moods on state anxiety and depression. It is also important to determine whether the identified impacts of happy/sad mood on state anxiety and depression can be replicated, especially since previous work has shown that inducing state depression using velten and musical procedures does not necessarily induce general negative mood (Clark, 1983). As such, additional research exploring the impact of WIM manipulations on state – in addition to trait -- anxiety and depression is also warranted.

Mechanisms for potential overlap between WIM and mental illness

Associative responding

The first means by which WIM and mental illness may be confounded is via associative responding. That is, it is possible that how people feel in the moment might impact the way in which they respond to questions about their mental illness due to increased availability of related associations. According to the Associative Network Theory, negative WIM makes negatively valanced information more salient whilst positive WIM makes positively valanced information more salient (Bower, 1981). This is because they activate a network of similar associations in the brain. Anxiety questionnaires ask people to report on the extent to which they experience different manifestations of anxiety, e.g., worrying about future events. Somebody who is feeling anxious before completing this questionnaire may be much more likely to recall times in which they felt anxious in the past since their emotional state is already priming them with specific associations of anxiety. Similarly, somebody who is feeling sad before completing a depression questionnaire that asks them about how often they tend to feel sad, is likely to respond in a heightened fashion. It is possible, therefore, that particularly in cases where people's emotional state is congruent with the symptomology being measured, mental illness scores will be artificially inflated.

Empirical overlap

Another means by which WIM and mental illness measures may be confounded is that affective states may interact with the wording of mental illness questionnaires to prime responses, resulting in the miscategorising of anxiety and depression. Constructivist emotion theories conceptualise moods as affective responses constructed in the moment by associated memories that are activated by the immediate environment (Gendron et al., 2018). These theories diverge from classical theories of emotion (Ekman, 1999), which consider emotions as having distinct, identifiable categories, and have gained increasing prevalence within emotion literature in recent years. According to this view, affective states are predictive rather than reactive. The brain uses current environmental input and relates this to similar past affective experiences to form an internal model (Barrett, 2017). This model generates affective experiences that help people to cope with current environmental demands and is frequently updated when new environmental input and associations are experienced.

As previously mentioned, according to the circumplex model, affective states activate two underlying neural pathways of valence and arousal (Russell, 1980). Valence determines how pleasant or unpleasant a person feels whilst arousal determines how alert and activated a person feels. Since negative affect is common to both anxiety and depression it follows that someone experiencing a generally negative affective state might be more likely to be primed by questionnaire wording that functions to help them interpret that negative state in either direction: anxious (negative high arousal) or depressed (negative low arousal). For example, somebody in a state of negative high arousal might feel more anxious when looking at a series of questions in which anxiety is the frame, e.g. “how anxious do you feel?”. Yet, if exposed to a series of questions in which depression is the frame, e.g. “how sad do you feel?” they might just as easily be primed in the opposite direction, interpreting their emotions as synched with those of depressive symptomology, due to the shared negative valence. If true, this may help to explain the finding that, whilst anxiety and depression are typically referred to as different constructs, various studies have demonstrated that they fail to show clear differentiating patterns (Eysenk & Fajkowska, 2018; Jamowski et al., 2019).

Affirming that current mental illness measures are not subject to empirical conflation is important since an inability to do so may contribute to theoretical errors in psychological research whereby certain behaviours and cognitions are thought to be associated with the wrong conditions. Misguided theory could lead to misdiagnosis of mental health conditions, increasing the likelihood of inefficient interventions targeting the wrong mechanisms. Demonstrating that anxiety and depression measures are robust to such effects can help to increase confidence in their validity and their utility, contributing to more robust theory on the interrelationship between these constructs (Watson et al., 2011).

2. The present work

To explore the separability of WIM and mental illness two online studies were conducted. Participants in both studies completed an autobiographical mood induction task followed by trait anxiety and depression questionnaires (study 1) and both trait and state anxiety and depression questionnaires (study 2). Autobiographical WIM induction relies on participants recalling past emotional experiences to evoke mood in a naturalistic manner (Prkachin et al., 1999) and has been shown to activate similar physiological responses to actual mood (Siedlecka & Denson, 2019). The WIM inductions targeted positively valanced emotional states that differed in arousal (happy – high arousal, relaxed – low arousal) and negatively valanced emotional states that differed in arousal (anxious – high arousal, sad – low arousal). Both studies were pre-registered[3].

Transparency and openness

This research provides appropriate citation for data and materials. For study one and two, the data and code for the main analysis can be found on the Open Science

[3] The study hypotheses were preregistered through the Open Science Framework (OSF; https://osf.io/stqbn, https://osf.io/rqmsx).

Foundation website[4]. The data cleaning code is not publicly available due to the fact it contains personal data, however, it can be sent over on request with an NDA. Study materials are clearly listed, and key indices (dependent variable scales) are written out in the Appendix for easy access. This research adheres to APA Design and Analysis Transparency Standards and it has been pre-registered. Study one was approved by the university board (unnamed as per masked review guidelines but can be provided at later date). Study two was approved by independent ethical regulators.

3. Study 1

Methods

Participants

There were 1081 UK English speaking participants that entered the study based on the following inclusion criteria: must be consenting adults >18 years old and score <11 on the Patient Health Questionnaire 8 measure of depression and fit within the 50:50 male/female sex assigned at birth quota. Of these 382 were excluded based on the following exclusion criteria: those who finished the survey in less than 400 seconds or more than 2250 seconds (these figures were determined using the duration histogram in the Appendix, S.1). A further 36 participants who failed to engage in the mood induction task appropriately were removed. This left 642 participants. A power analysis conducted using “pwr” package in R for a one-way between people ANOVA suggested that a sample size of 128 per group would allow for an 87% chance of detecting a small (f=.15) effect size.

[4] The Open Science Framework (OSF; https://osf.io/stqbn, https://osf.io/rqmsx).

Gender and age were distributed fairly evenly across their respective groups (Appendix Tables 1-2). Participants had a median anxiety score of 42 and a median depression score of 8. 397 participants had anxiety scores above 38 indicating mild, moderate, or severe anxiety (62% of the total sample). 196 participants had depression scores above 13 indicating mild, moderate, or severe depression (31% of the total sample). All participants were recruited via Pure Profile survey recruitment company and received £5 for participation.

Measures

Depression screening (study 1 only). The *Patient Health Questionnaire-8* was used to screen participants out of the study (Kroenke & Spitzer, 2009). Eight items assess participants' perception of their depressive symptoms by asking "Over the last two weeks, how often have you been bothered by each of the following problems" – an example item is, "feeling down, depressed, or hopeless". Scoring ranges from 0 (no depression) to 24 (severe depression).

General wellbeing. The *Office of National Statistics personal wellbeing questions* (ONS, 2018). These items were scored on a scale of 0-10 where 0 denotes "not at all" and 10 denotes "completely". The 4 items are listed below:

> Overall, how satisfied are you with your life nowadays?
>
> Overall, to what extent do you feel the things you do in your life are worthwhile?
>
> Overall, how happy did you feel yesterday?
>
> Overall, how anxious did you feel yesterday?

Personality. The *Brief Ten Item Personality Index* (Gosling et al., 2003) was included in this study since personality has been shown to impact the success of mood induction (Larsen & Ketelaar, 1989). This 10-item scale asks people to self-report the extent to which a list of characteristics apply to them, such as "Extroverted, enthusiastic" for example. The five personality traits (extroversion, agreeableness, conscientiousness, emotional stability, and openness to experiences) are calculated by adding the scores of two items each. Scores are recorded on a 0-7 Likert scale where 0 denotes “disagree strongly” and 7 denotes “agree strongly”.

Social desirability. The short 13-item Marlowe-Crowne Social Desirability Scale (Andrews & Meyer, 2003) was used to control for potential demand effects in self-reported mood. Using this scale participants must indicate whether several socially desirable statements are true or false as they pertain to them. An example item is “It is sometimes hard for me to go on with my work if I am not encouraged”.

WIM (valence and arousal). The Affective Slider (Betella & Verschure, 2016) allows for the relative assessment of negative (anxious, sad) and positive (happy, relaxed) moods on a comparative scale. This is important when trying to determine which properties of moods might be linked to changes in mental illness reports since different moods may evoke different combinations of valence and arousal for different people (Mills & D’Mello, 2014). The measure consists of two separate sliders, one pertaining to level of arousal and another pertaining to level of pleasure right now (exact wording can be seen in Appendix Figure 1).

IV: Mood induction

Mood was induced using the autobiographical memory recall technique (Krauth-Gruber & Ric, 2000). This is one of the most effective, simple to conduct and commonly used mood induction methods (Jallais & Gilet, 2010). It allows for the induction of different moods online in a consistent manner across conditions. It has also been shown to reliably impact self-reported mood as well as non-subjective indicators of mood such as physiological responses and cognitive processes (Robinson & Sahakian, 2009; Robinson et al., 2012). The fact that this mood induction relies on recall of real-life affective experiences was also considered preferable for external validity. The task requires participants to write down a time when they felt relaxed, sad, happy or anxious in as much detail as possible. For the neutral condition, participants were asked to recall an ordinary event from their past consistent with previous research (Sidlecka et al., 2015). Participants were asked to spend 4 minutes on this task. Online mood induction has been shown to be effective for most affective states, finding similar effect sizes to mood inductions in laboratory settings (Ferrer & Grenen, 2015; Lench, Flores & Bench, 2011).

WIM (valence and arousal). Repeat of above WIM measure.

DV: Trait depression. The *Beck Depression Inventory* (BDI – Beck et al., 1961) is a widely used tool for measuring depression (Shafer, 2006). This is a 21-item measure which has been shown to have an alpha coefficient of above .90 in many populations (Beck et al., 1988). In each item participants must select one out of a list of 4 statements that most applies to them, e.g. "I do not feel sad; I feel sad; I am sad all of the time and I can't snap out of it; I am so sad and unhappy that I can't stand it". Scores range from 0-63: 0-13 = minimal depression, 14-19 = mild depression, 20-28 = moderate depression and 29-63 = severe depression.

DV: Trait anxiety. The trait items from the *State-Trait Anxiety Inventory* (STAI – Spielberger, 1980). This 20-item questionnaire asks participants to indicate how they "generally feel" and consists of items such as: "I worry too much over something that doesn't really matter". Items are rated on a 0 (almost never) to 3 (almost always) scale. There is considerable evidence supporting the construct and concurrent validity of the scale (Spielberger, 1989). Scores range from 20 to 80: 20-37 = no or low anxiety, 38-44 = moderate anxiety and 45-80 = high anxiety).

*The presentation order of the depression and anxiety questionnaires above was randomised.

Procedure

An overview of the procedures for study 1 and 2 is depicted in Figure 1 below. More details on study 2 can be found in the study 2 section.

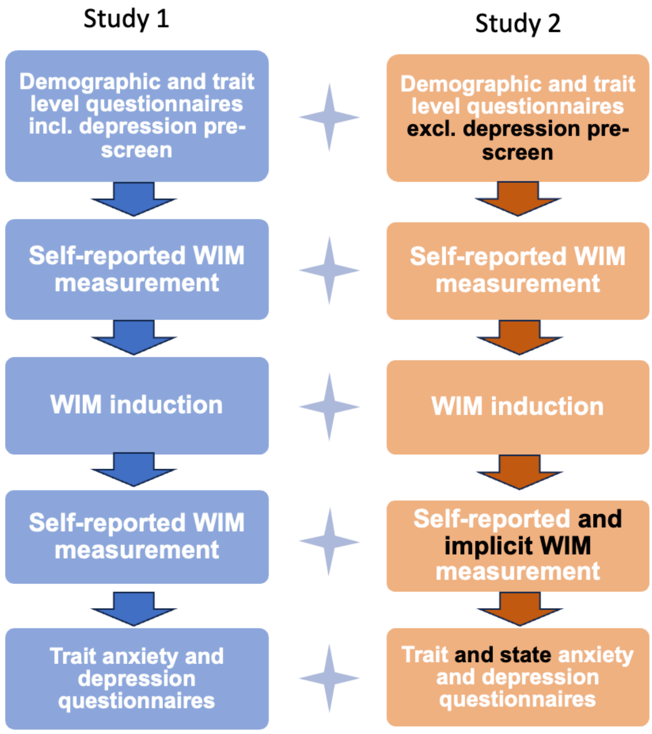


Figure 1. *A schematic diagram depicting an overview of the procedures for study 1 and 2. The main differences between the two studies are marked in black.*

Participants first completed demographic questions (age and gender) followed by trait level questionnaires (general wellbeing, personality, and social desirability measures). They were then shown a short video of a bubble intended to neutralise mood in preparation for the WIM induction. This was followed by self-reports of valence and arousal, then the autobiographical mood induction task. During the WIM induction participants were given 4 minutes to recall a time in which they felt either happy, sad, relaxed, or anxious and write about it in detail, in line with previous research (Becker & Leineger, 2011). Those in the neutral condition were asked to recall and write about a time in which they experienced an ordinary event. After the WIM induction, participants were asked to restate their valence and arousal levels before completing the anxiety and depression questionnaires (the presentation order of the anxiety and depression questionnaires was randomised across participants).

Predictions and associated analysis

Assuming an impact of WIM on anxiety and depression questionnaire responses:

> **H1**: Positive valence (happy and relaxed) will predict decreases in anxiety/depression relative to more negative valence (sad, anxious and neutral moods)

If due to associative responding, these effects will be greater under conditions where arousal levels are matched with the questionnaire:

> **H2:** High arousal, negative valence (anxious) will predict greater increases in anxiety than depression relative to low arousal, negative valence (sad)
>
> **H3:** Low arousal, negative valence (sad) will predict greater increases in depression than anxiety relative to high arousal, negative valence (anxiety)

If due to anxiety and depression questionnaire wording priming responses in either direction, arousal congruence shouldn't matter:

> **H4:** Higher arousal, negative valence (anxious) will predict **similar increases in both** anxiety and depression scores than lower arousal, negative valence (sad)
>
> **H5:** Lower arousal, negative valence (sad) will predict **similar increases in both** anxiety and depression scores than higher arousal, negative valence (anxious)

To check whether participant's WIM differed before the induction, two separate ANOVAs were conducted with pre-valence/arousal levels as the dependent variable and the five experimental groups (happy, sad, relaxed, anxious, and neutral) as independent variables. To check the success of the mood induction, two ANOVAs were conducted with valence/arousal delta as the dependent variable (pre-valence/arousal minus post-valence/arousal) and the five experimental groups as independent variables.

For robustness, and to check whether people might have responded differently to the WIM inductions based on their baseline affective traits/states which might compromise findings, additional linear regressions were conducted on valence and arousal delta with the following predictors included: social desirability; personality: extroversion, openness, emotional stability, agreeableness, and conscientiousness; wellbeing; initial depression as measured by the PHQ8 (study 1 only); age and gender. Linear regressions were selected here since they tend to be preferable when using continuous variables as predictors (Agresti, 2018).

An analysis of variance (ANOVA) was used to compare the effects of the five different WIM induction groups on trait anxiety and depression, respectively. The neutral group was selected as the reference group for these analyses. ANOVA's were chosen since they are usually preferable to regressions when predicting the impact of categorical variables on a given outcome (Agresti, 2018). The same analyses were then repeated but this time focusing only on those in either the top (scores <13) or bottom (scores >25) quartile of wellbeing scores. This was to check whether the mental illness scores of people with very high or low wellbeing were affected by mood more or less than the average person.

For completeness, linear regressions exploring the impact of raw valence and arousal delta scores on trait anxiety and depression scores were also conducted. Regressions were again chosen here since they are generally preferable when exploring the impact of continuous variables on a given outcome (Agresti, 2018). Since regression analysis allows for it, these analyses also incorporated all additional variables measured in the study (agreeableness, extraversion, emotional stability, openness, social desirability, and wellbeing).

False Discovery Rate (FDR) corrections were applied to all p values (one set for the ANOVAs and another set for the regressions) to reduce the likelihood of false positive findings considering that multiple tests were run (Benjamini & Hochberg, 1995). Cohen's *d* was computed to assess the magnitude of the significant effects.

Results

WIM induction

ANOVAs confirmed no significant differences in pre-valence/arousal levels between groups (pre valence: F (4, 637) = 0.975, p = 0.421; pre arousal: F (4, 637) = 0.658, p = 0.621). Levene's tests on pre mood induction valence and arousal scores revealed no evidence that the variance across groups was statistically significantly different.

Average valence change scores (pre - post self-reported valence) indicated that both happy (M = 7.72, SD = 25.5, *t* = -5.31, p = <0.001) and relaxed (M = 5.40, SD = 19.4, *t* = -3.74, p = <0.001) moods increased valence (higher positive scores indicate higher positive valence increase). Neutral mood generated a small decrease in valence (M = -3.84, SD = 20.3, *t* = -1.07, p = 0.28). Anxious mood generated a larger decrease in valence (M = -10.7, SD = 26.4, *t* = 4.32, p = <0.001) than neutral mood, and sad mood generated the largest decrease in valence (M = -17.2, SD = 25, *t* = 7.84, p = <0.001).

An ANOVA with the five experimental groups as independent variables and valence delta as the dependent variable revealed a significant difference in valence between groups (F (4, 637) = 25.62, p <0.001***), confirming the success of the WIM induction on self-reported valence. Planned comparisons (Appendix Table 3) showed that the following groups differed in self-reported valence as expected: happy/anxious, relaxed/anxious, neutral/happy, sad/happy, relaxed/neutral, sad/neutral, sad/relaxed. Sad/anxious, neutral/anxious, and relaxed/happy did not differ in valence. However, given that the apriori Levene's test for valence revealed a significant difference in valence change across mood induction conditions ($f=4.050$, $p<0.003$**) a non-parametric Kruskal-Wallis test was run to account for the possible impact of non-equal variance on this result. We note that the significant impact of mood induction conditions on valence delta remained.

In terms of arousal, as expected happy mood increased arousal ($M = 7.65$, $SD = 24.6$) and sad mood decreased arousal ($M = -7.84$, $SD = 21.5$). However, counter to expectations anxiety had a negative impact on self-reported arousal ($M = -1.81$, $SD = 23.4$) and relaxed had a positive impact on self-reported arousal ($M = 5.44$, $SD = 23$). Since all positive emotions were assessed as higher in arousal and all negative moods were assessed as lower in arousal it may be that participants interpreted high arousal as a positive and low arousal as negative. This possibility is considered when drawing conclusions.

An ANOVA with the five experimental groups as independent variables and arousal delta as the dependent variable revealed significant differences across groups ($F(4, 637) = 9.171$, $p < 0.001$***). Levene's test confirmed no significant differences in variance for arousal delta across conditions. Planned comparisons (Appendix Table 4) showed that the following groups differed in self-reported arousal: happy/anxious, sad/happy, sad/neutral, and sad/relaxed. The difference between relaxed and anxious was approaching significance. Though as stated earlier, except for sad/relaxed and sad/neutral, we interpret these findings with caution since they are at odds with what theory would predict. Namely, anxious should be a more highly aroused mood than relaxed and it should be similar in arousal levels to happy.

For robustness, social desirability was tested as a potential predictor of valence and arousal delta alongside other continuous covariates in a linear regression (See Appendix Tables 5 and 6). A simple linear regression showed no significant relationship between any of these variables with valence/arousal delta. The linear regression ran on valence and arousal delta including additional control variables (social desirability; personality: extroversion, openness, emotional stability, agreeableness, and conscientiousness; wellbeing; initial depression as measured by the PHQ8; age and gender) revealed none of these as being significant predictors.

The impact of WIM on trait anxiety and depression reports

Descriptive statistics for trait anxiety and depression reports across groups are reported in Tables 1 and 2.

Table 1. Mean trait anxiety scores by group

Group	Count	STAI score	Standard deviation
Anxious	130	41.4	12.0
Happy	138	41.7	11.4
Neutral	128	42.4	11.2
Relaxed	126	40.6	11.9
Sad	120	41.8	9.61

Table 2. Mean trait depression scores by group

Group	Count	BDI score	Standard deviation
Anxious	130	9.78	8.48
Happy	138	9.01	8.31
Neutral	128	10.1	8.44
Relaxed	126	8.67	8.36
Sad	120	9.5	8.15

ANOVAs on both trait anxiety and depression scales across groups confirmed no statistically significant differences (anxiety: $F(4, 637) = 0.443$, $p = 0.778$; depression: $F(4, 637) = 0.583$, $p = 0.675$). The assumption of homogeneity of variances was tested and satisfied based on Levene's F test for depression ($F(4, 637) = 0.170$, $p = 0.953$). However, this assumption was not satisfied for trait anxiety ($F(4, 637) = 2.839$, $p = 0.024$*). Therefore, a Kruskal-Wallis test was used to assess whether trait anxiety scores differed according to the different mood induction conditions. This test also revealed a non-significant result (H(4) = 2.036, p = 0.729). A further Kruskal-Wallis test was conducted as a robustness check for the impact of mood induction on trait depression scores, given that responses to this variable were negatively skewed. These results also remained non-significant (H(4) = 3.352, p = 0.501).

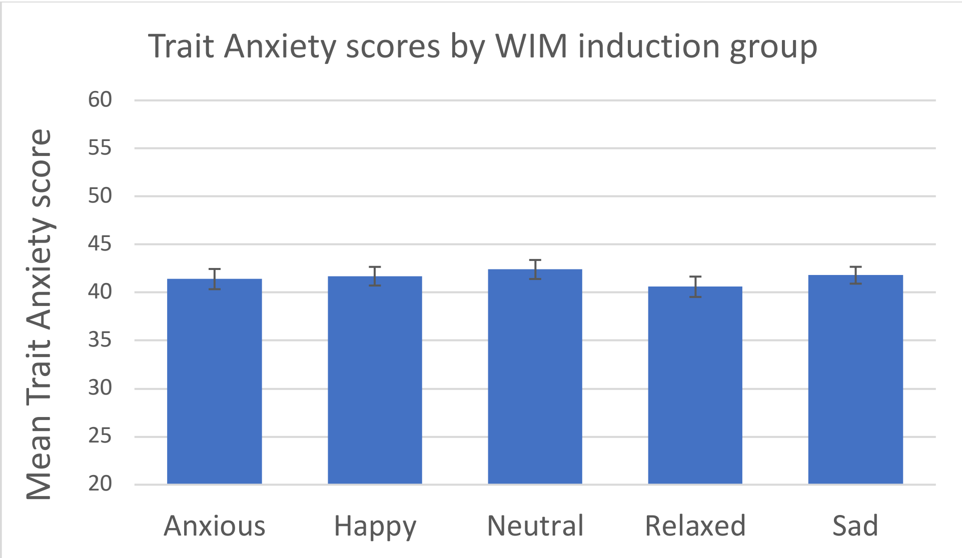


Figure 2. *Mean trait anxiety scores and standard errors per WIM induction group.*

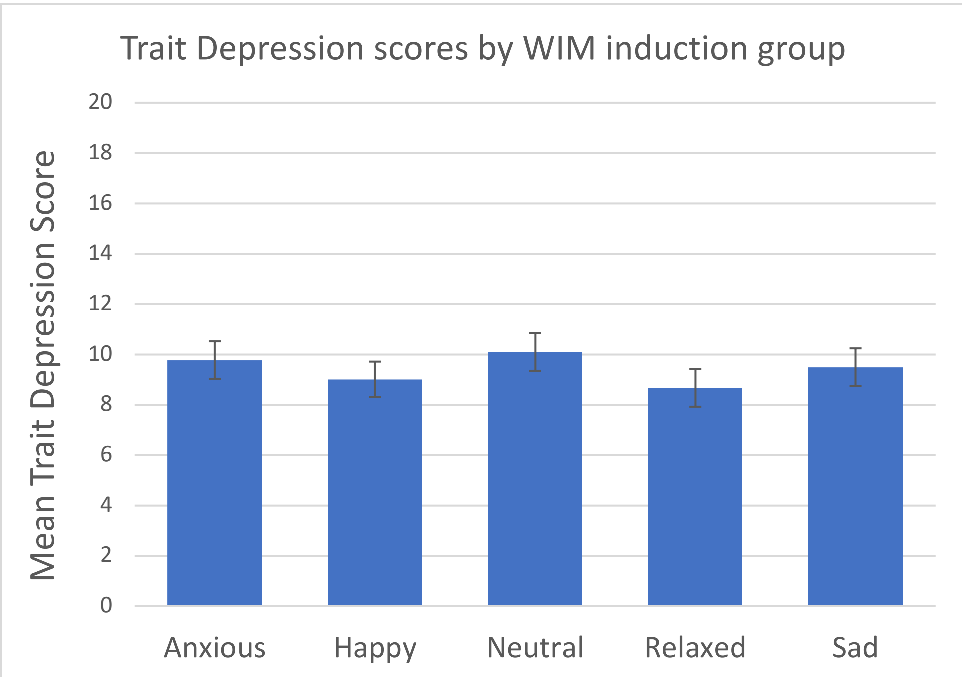


Figure 3. *Mean trait depression scores and standard errors per WIM induction group.*

The results from the ANOVAs that were run on high and low wellbeing groups respectively, remained consistent, revealing non-significant impacts of WIM across groups for both trait anxiety (high wellbeing: $F(4, 160) = 0.246$, $p = 0.912$; low wellbeing: $F(4, 441) = 0.776$, $p = 0.541$) and trait depression (high wellbeing: $F(4, 160) = 0.656$, $p = 0.624$; low wellbeing: $F(4, 446) = 0.549$, $p = 0.700$). The linear regression models exploring whether raw valence and arousal delta scores predicted trait anxiety (valence delta: $b = -0.009$, SE = 0.010, $p = 0.388$; arousal delta: $b = -0.010$, SE = 0.011, $p = 0.380$) and depression (valence delta: $b = 0.000$, SE = 0.009, $p = 0.964$; arousal delta: $b = -0.031$, SE = 0.000, $p = 0.076$) also revealed non-significant results. See Appendix Tables 7-8 for full regression tables.

Study 1 discussion

Study 1 found that inducing WIM across discrete mood groups (happy/sad/relaxed/anxious/neutral) did not impact scores on trait anxiety or depression scales. This effect remained when examining individuals who scored relatively low or high on initial wellbeing, suggesting that how people felt to begin with did not make them any more likely to change their mental illness scores in response to shifts in WIM. It also remained when examining the impact of raw valence and arousal change scores on anxiety and depression. Given the null impact of WIM induction on trait anxiety and depression, none of the study hypotheses 1-6 were supported.

These findings lend support to theory suggesting that WIM is a separate and therefore distinguishable construct from trait anxiety and depression. They also affirm the validity of anxiety and depression scales, as measures that are not easily influenced by transient factors such as WIM.

4. Study 2

Methods

Participants

There were 643 UK English speaking participants that entered the study based on the following inclusion criteria: must be consenting adults >18 years old and fit within the nationally representative quotas set on gender, ethnicity, and age. Of these 643 were excluded based on the following criteria: those who finished the survey in less than 600 seconds (this figure was determined using the duration histogram, see Appendix Figure 5 for more details). A further 49 participants who were either unable to remember a time in which they had experienced the target mood or failed to engage in a mood recollection (e.g. by answering "n/a") were removed. This left 565 participants which amounted to roughly 113 per condition. A power analysis conducted using "pwr" package in R for a one-way between people ANOVA confirmed that a sample size of 113 per group will allow for an 82% chance of detecting a small (f =.15) effect size at significance level 0.05.

The sample were nationally representative with respect to gender, age and ethnicity (see Appendix Tables 9-11). Trait anxiety and depression scores were slightly higher in this sample (Median Anxiety = 44, Median depression = 9) than in study 1 (Median Anxiety = 42, Median depression = 8). 375 participants had trait anxiety scores above 38 indicating mild, moderate or severe anxiety (66% of the total sample). 224 participants had trait depression scores above 13 indicating mild, moderate or severe depression (40% of the total sample). All participants were recruited via Pure Profile survey recruitment company and received £5 for participation.

Measures

Most measures remained the same as in study 1 for study 2. However, the depression screening questionnaire was removed, and state anxiety and depression questionnaires were included. An implicit mood measure was also included. See additions below.

Implicit WIM measure. Given that previous work has shown negative WIM induction to impact hedonic judgement of mildly positive and negative words (Baddeley et al., 2012), we incorporated word evaluation questions as a more implicit means to monitor the success of the WIM induction in study 2. Participants were asked to rate two mildly affective words taken from the Affective Norms for English Words (ANEW database, Bradley & Lang, 1999) from negative (1) to positive (9) and from non-arousing/non-stimulating (1) to arousing/stimulating (9). The words chosen were false (mildly negative and mildly de-arousing, M valence = 3.27, M arousal = 3.43), and body (mildly positive and mildly arousing, M valence = 5.55, M arousal = 5.52).

State Anxiety and Depression. State anxiety and depression were measured using the State-Trait-Anxiety-Depression-Index (Laux et al., 2018). The state part, which this study is concerned with, consists of 10 items assessing state anxiety which is made up of 5-item emotionality (e.g., "I feel uneasy") and 5-item worry ("I worry about my situation") subcomponents and 10 items assessing depression which is made up of 5-item euthymia ("I am glad" reverse scored) and 5-item dysthymia ("I am dejected") subcomponents. The euthymia items are reverse scored to yield anhedonia scores. Responses are measures using four response options ranging from 1 (almost never) to 4 (almost always). This relatively new measure has demonstrated good reliability and validity (Brahler et al., 2002). Scores range between 10 and 40 for state anxiety and depression respectively.

Procedure

Study 2 was a direct replication of study 1 with a few additions. Study 2 included two state measures of anxiety and depression in addition to the trait measures. This was to account for temporal differences between the instructions for the trait (refers to general feelings) and WIM measures (refers to feelings right now). It also included a word evaluation task as an implicit measure of WIM. Study 2 also included a more encompassing sample by not excluding depressed participants from taking part.

Predictions and associated analysis

This remained the same as in study 1. The analysis included two additional linear regressions on state anxiety and depression. As per the trait level regressions, additional measured variables were added as predictor variables for robustness (extraversion, openness, emotional stability, agreeableness, wellbeing, social desirability, age, and gender).

Results

Descriptive statistics and WIM induction check

ANOVAs with pre-valence/arousal levels as the dependent variable and the five experimental groups (happy, sad, relaxed, anxious, and neutral) as independent variables revealed no significant differences in pre-valence/arousal levels between groups (pre valence: $F(4, 560) = 0.483$ $p=0.748$; pre arousal: $F(4, 560) = 0.269$, $p=0.898$).

Average valence change scores (pre - post self-reported valence) indicated that both happy ($M = 10.3$, $SD = 20.5$) and relaxed ($M = 7.43$, $SD = 20.5$) moods increased valence (higher positive scores indicate higher positive valence). Neutral mood generated a small increase in valence ($M = 1.93$, $SD = 19.9$). Both sad ($M = -20.9$, $SD = 28.2$) and anxious ($M = -11.2$, $SD = 27.6$) moods generated a decrease in valence. An ANOVA with mood predicting valence delta revealed a significant difference in valence between conditions ($F(4, 560) = 38.87$, $p < 0.001$***), confirming the success of the mood induction on self-reported valence. Planned comparisons (Appendix Table 12) showed that the following groups differed in self-reported valence as expected: happy/anxious, happy/neutral, relaxed/anxious, sad/happy, relaxed/neutral, sad/neutral, sad/relaxed. As expected, sad/anxious and relaxed/happy did not differ in valence. There was no significant difference in arousal change between anxious and neutral conditions. These results were largely consistent with study 1.

In terms of arousal, as expected the happy mood induction increased arousal ($M = 9.53$, $SD = 20.5$) and the sad mood induction decreased arousal ($M = -6.59$, $SD = 28.2$). However, counter to expectations anxious mood ($M = -2.51$, $SD = 27.6$) had a slightly negative impact on self-reported arousal and both neutral ($M = 5.07$, $SD = 19.9$) and relaxed mood ($M = 6.68$, $SD = 20.5$) had a positive impact on self-reported arousal. An ANOVA with mood predicting arousal delta revealed significant differences across conditions ($F(4, 560) = 8.1$, $p < 0.001$***). Planned comparisons (Appendix Table 13) showed that the following groups differed in self-reported arousal: happy/anxious, sad/happy, sad/neutral, and sad/relaxed. There was only a near significant difference between relaxed/anxious and like study 1 it was in the opposite direction than expected with relaxed mood increasing arousal and anxious mood decreasing arousal. Neutral/anxious, sad/anxious, neutral/happy, relaxed/happy and relaxed/neutral did not differ in arousal.

The linear regression ran on valence and arousal delta including additional control variables (social desirability; personality: extroversion, openness, emotional stability, agreeableness, and conscientiousness; wellbeing; age and gender) revealed none of these as significant predictors (Appendix Tables 14-15).

Implicit WIM

Mean self-reported positive word evaluations ranged from 5.87 to 6.29 with happy being most positive and neutral/relaxed being least positive (Appendix Table 16). Mean self-reported negative word evaluations ranged from 2.65 to 3.21 with sad being most negative and relaxed being least negative (Appendix Table 17).

The ANOVA conducted on implicit negative word evaluation scores revealed a close to significant impact ($F(4, 560) = 2.073$, $p = 0.083$). The Sad-Relaxed group difference had the lowest p value out of the pairwise comparisons ($p = 0.181$) with sad mood resulting in more negative evaluations than relaxed mood. However, there was no clear pattern here since happy mood rated negative words similarly to sad mood. The ANOVA conducted on implicit positive word evaluation scores also revealed a non-significant impact ($F(4, 560) = 1.015$, $p = 0.399$).

The impact of WIM on trait anxiety and depression scores

Descriptive statistics for trait anxiety and depression reports across groups are displayed in Tables 3 and 4. Across all mood induction conditions, trait anxiety and depression reports are very similar.

Table 5. Mean trait anxiety scores by group

Group	Count	STAI score	Standard deviation
Anxious	113	46.1	13.7
Happy	112	42.9	12.0
Neutral	121	42.8	12.4
Relaxed	107	43.3	13.6
Sad	112	44.6	11.9

Table 6. Mean trait depression scores by group

Group	Count	BDI score	Standard deviation
Anxious	113	13.6	11.9
Happy	112	10.8	9.18
Neutral	121	10.8	9.52
Relaxed	107	12.0	11.5
Sad	112	10.8	8.56

ANOVAs on both trait anxiety and depression scales across groups confirmed no statistically significant differences (anxiety: $F(4, 560) = 1.378$, $p = 0.24$; depression: $F(4, 560) = 1.673$, $p = 0.155$). The assumption of homogeneity of variances was tested and satisfied based on Levene's F test for trait anxiety ($F(4, 560) = 1.078$, $p = 0.367$). However, this assumption was not satisfied for trait depression ($F(4, 560) = 3.835$, $p = <0.01$**) Therefore, a Kruskal-Wallis test was used to assess whether trait depression scores differed according to the different mood induction conditions. This test also revealed a non-significant result ($H(4) = 3.19$, $p = 0.527$).

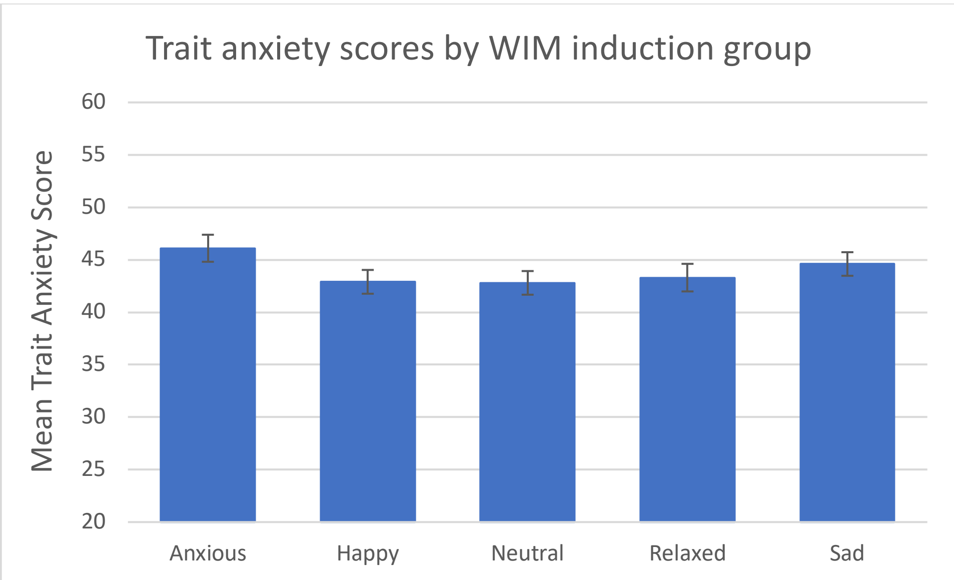


Figure 4. *Mean trait anxiety scores and standard errors per WIM induction group.*

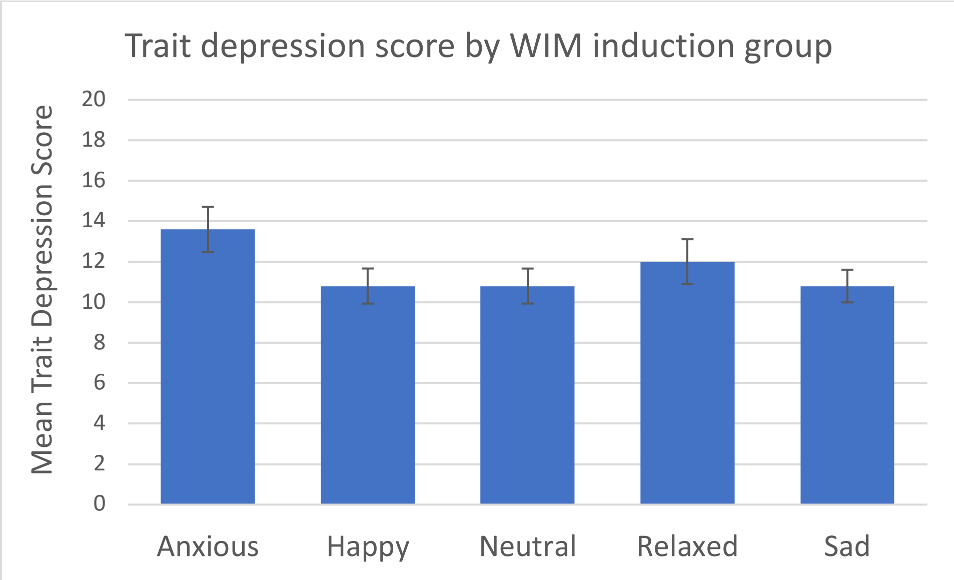


Figure 5. *Mean trait depression scores and standard errors per WIM induction group.*

The results from ANOVAs conducted on those in either the top (scores <13) or bottom (scores >25) quartile of wellbeing scores remained consistent across groups for trait anxiety (high wellbeing: $F(4, 157) = 1.469$, $p = 0.214$; low wellbeing: $F(4, 367) = 2.1$, $p = 0.080$). For trait depression the low wellbeing group remained non-significant (low wellbeing: $F(4, 367) = 0.588$, $p = 0.672$) but the high wellbeing group did reach significance (high wellbeing: $F(4, 157) = 2.525$, $p = 0.043$*). Pairwise comparisons revealed that for those scoring high on wellbeing, sad mood increases trait depression scores relative to relaxed mood ($p = 0.038$*). However, this effect disappeared once multiple comparisons were accounted for using the FDR method.

Further linear regression models with valence and arousal delta predicting trait anxiety revealed no significant impacts on anxiety (valence delta: $b = -0.010$, SE = 0.011, $p = 0.357$; arousal delta: $b = -0.003$, SE = 0.011, $p = 0.786$). For the linear regressions predicting trait depression, a non-significant impact was found for valence delta, but a significant negative impact was found for arousal delta (valence delta: $b = 0.022$, SE = 0.012, $p = 0.073$; arousal delta: $b = -0.031$, SE = 0.012, $p = 0.011$*). With every one unit increase in arousal there was a 0.031 decrease in trait depression scores. The Cohen's *d* computed effect size was 0.106, constituting a small effect. This effect remained significant post FDR corrections. Full regression tables can be found in the Appendix Tables 18-19.

The impact of WIM on state anxiety and depression scores

Descriptive statistics for state anxiety and depression reports across groups are displayed in Tables 6 and 7. Across all WIM induction conditions, state anxiety and depression reports are very similar.

Table 6. Mean state anxiety scores by group

Group	Count	STAI score	Standard deviation
Anxious	113	20.1	7.92
Happy	112	17.7	6.39
Neutral	121	18.6	7.29
Relaxed	107	18.2	7.30
Sad	112	18.2	6.32

Table 7. Mean state depression scores by group

Group	Count	BDI score	Standard deviation
Anxious	113	22.2	4.05
Happy	112	19.4	3.25
Neutral	121	20.8	4.05
Relaxed	107	21	4.37
Sad	112	21.7	3.93

The ANOVA on state anxiety confirmed no statistically significant differences ($F(4, 560) = 1.882$, $p = 0.112$). The ANOVA on state depression revealed a statistically significant difference ($F(4, 560) = 2.723$, $p = 0.029$*). Pairwise comparisons revealed that those in the happy condition scored significantly lower on depression than those in the anxious condition ($p = 0.020$*, see Appendix Table 20). However, this difference became insignificant post FDR corrections. The assumption of homogeneity of variances was tested and satisfied for this analysis based on Levene's F test for state anxiety ($F(4, 560) = 1.529$, $p = 0.192$) and state depression ($F(4, 560) = 0.431$, $p = 0.787$).

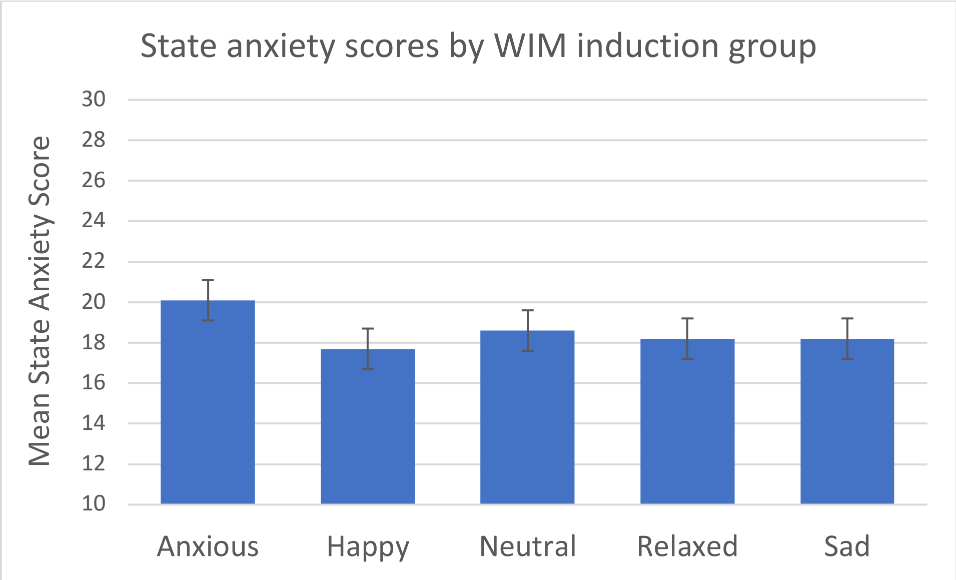


Figure 6. *Mean state anxiety scores and standard errors per WIM induction group.*

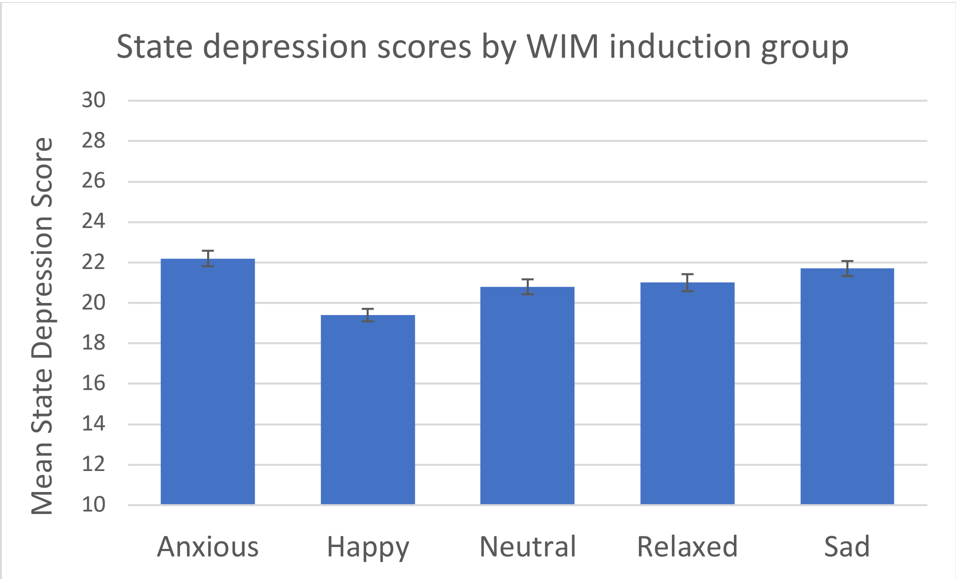


Figure 7. *Mean state depression scores and standard errors per WIM induction group.*

The same analyses were conducted on those in either the top (scores <13) or bottom (scores >25) quartile of wellbeing scores. The results remained consistent across groups for trait anxiety (high wellbeing: $F(4, 157) = 1.449$, $p = 0.22$; low wellbeing: $F(4, 367) = 0.747$, $p = 0.561$). For trait depression the low and high wellbeing group remained non-significant (low wellbeing: $F(4, 367) = 2.057$, $p = 0.082$; high wellbeing: $F(4, 367) = 1.665$, $p = 0.158$).

Further linear regression models with valence and arousal delta predicting state anxiety revealed no significant impacts (valence delta: b = -0.002, SE = 0.009, p = 0.841; arousal delta: b = -0.010, SE = 0.009, p = 0.303). For the linear regressions predicting state depression, however, a significant impact was found for valence delta (b = -0.020, SE = 0.007, p =0.005**) and for arousal delta (b = -0.015, SE = 0.007, p = 0.045*). For every one unit increase in positive valence state depression scores decreased by 0.020 and for every one unit increase in arousal state depression scores decreased by 0.015. Only the significant effect of arousal delta on state depression remained post FDR corrections. The Cohen's d computed effect size for this impact was 0.119, constituting a small effect. Full regression tables can be found in the Appendix Tables 22-22.

Discussion

Like study 1, study 2 also found that inducing WIM across discrete mood groups (happy/sad/relaxed/anxious/neutral) did not impact scores on trait anxiety or depression scales. This non-significant effect remained when examining individuals who scored relatively low or high on initial wellbeing once multiple comparisons had been accounted for. Therefore, how people felt to begin with did not make them any more likely to change their mental illness scores in response to shifts in WIM.

These non-significant effects remained when examining the impact of raw valence and arousal change scores on trait anxiety. However, when examining the impact of raw valence and arousal change scores on trait depression, a significant negative impact of arousal change was observed, which remained post FDR corrections. Here it was found that a one unit increase in arousal was associated with a 0.031 unit decrease in trait depression scores, constituting a small effect (Cohen's d = 0.106). This lends support to the arousal component of H3, "**Low arousal**, negative valence (sad) **will predict greater increases in depression than anxiety relative to high arousal**, negative valence (anxiety)".

With respect to the impact of discrete WIM conditions (happy/sad/relaxed/anxious/neutral) on state anxiety, no significant impacts of WIM were identified. This non-significant effect remained for those with relatively high and low wellbeing to begin with. Valence and arousal delta scores also revealed a non-significant impact on state anxiety scores. For state depression, there was also no significant impact of discrete WIM conditions (happy/sad/relaxed/anxious/neutral) identified in the ANOVAs, once multiple comparisons had been accounted for.

In the regression exploring the impact of raw valence/arousal delta scores on state depression, however, a significant impact of valence delta was detected. Specifically, it was found that post FDR corrections, higher valence scores were associated with lower depression scores. With every one unit increase in positive valence state depression scores decreased by 0.020 This constituted a small effect (Cohen's *d* = 0.119). This lends support for valence component of H1, "**Positive valence** (happy and relaxed) **will predict decreases in anxiety/depression relative to more negative valence** (sad, anxious and neutral moods)".

The WIM inductions were unsuccessful at changing people's positive/negative word evaluations. This suggests that either the WIM inductions were not strong enough to be picked up by these more subtle, implicit measures, or people were overstating the impact of the inductions in their self-reports due to experimental demand effects. The former is perhaps more likely given that no association between social desirability and self-reported mood was identified, which would have been expected if experimental demand effects were present. Moreover, many previous studies have correlated impacts of identical autobiographical inductions with physiological responses considered to be effective proxies for WIM.

Overall, the findings from study 2 have revealed some evidence for a small impact of WIM on trait and state depression and this impact is consistent with the Associative Network Theory.

5. **General discussion** (note this only relates to paper 2)

Are WIM and mental illness separable and therefore distinct constructs? The answer to this question is vital in helping us to affirm the measurement and theoretical grounding of these two constructs. Practically, it can also help to ensure the right interventions and diagnoses are given to the right people. Study 1 set out to explore this question by assessing the impact of WIM manipulations (happy, sad, relaxed, anxious, neutral) on trait anxiety and depression reports. Study 2 was a replication of study 1 with the additional inclusion of *state* anxiety and depression measures. Both studies identified no significant impact of discrete WIM conditions on trait (studies 1 and 2) and state (study 2 only) anxiety and depression. When assessing the impact of valence and arousal change scores on trait and state anxiety and depression, however, study 2 identified a significant impact of arousal change on trait depression, and valence change on state depression. Where significant impacts were identified, they were consistent with what might be predicted by Associative Network Theory, whereby elevated scores occurred under conditions where the WIM experience was congruent with the target questionnaire.

The fact that WIM manipulations do not alter responses to the Trait-Anxiety-Index (taken from STAI) and State-Anxiety-Inventory (taken from STADI) supports the theoretical assumption that WIM and anxiety are indeed separable and therefore distinct constructs. If these constructs had shared variance, we would have expected changes in WIM to reveal themselves in mental illness reports. Given the null impact of WIM on trait and state anxiety reports, these findings demonstrated no evidence of associative responding or empirical overlap, predicted in H1-5. One reason for this null result on trait/state anxiety may be that the cognitive associations recalled due to WIM are not similar enough to those recalled when responding to anxiety questionnaire items, and therefore do not influence those mental illness reports. In line with Associative Network Theory (Bower, 1981), if the associations generated by WIM were similar enough, we should see an impact of WIM on anxiety reports in conditions where WIM is congruent with anxiety (e.g. low valence or low valence and high arousal). From a constructivist perspective, it may be that anxious WIM related concepts constructed in the moment don't carry over to anxiety related concepts constructed in the moment. One way for future studies to test this would be to explore whether different free associations are generated when people are primed to think about anxious mood as compared with free associations generated when people are primed to think about anxiety symptomology.

The null impact of WIM on the Trait-Anxiety-Index is consistent with previous work showing that anxiety inductions, this time generated by communicating task failure, do not impact trait anxiety scores using the same trait measure (Gaudry et al., 1975). However, it is interesting that no such impact of WIM was identified for state anxiety, since past studies have revealed an impact here both in response to anxious WIM (Gaudry et al., 1975) and in response to happy and sad WIM inductions (Baker, 1993). This difference in result may be due to the anxiety manipulation in the present study not being strong enough. Alternatively, it may be that the state anxiety used in the present study is more robust to WIM manipulations than the State-Trait-Anxiety-Index, which was used in both studies where a direct impact was identified. If true, the STADI may be preferred when it comes to differentiating clinically problematic from non-clinically problematic WIM experiences as a more effective tool. Future work can seek to test this assumption by assigning one anxious induction group to the STADI and another to the STAI and comparing results.

Being able to distinguish between anxious mood and trait/state anxiety has important theoretical and practical implications. If these null findings continue to be replicated in future research, theories of WIM and anxiety can be clearer in describing WIM as a pathway to, rather than a component of, anxiety and vice versa. Given this, it should not be assumed that eliciting changes in one will necessarily elicit changes in the other. For example, feeling sad/happy in the moment may not directly translate to somebody experiencing more/less mental illness symptoms, though over the long-term this may be possible. Recent work highlighting the important role of frequent fluctuations in WIM over time in determining mental disorders, is a good example of how WIM can be conceived as a pathway to reducing mental illness over time (Patel et al., 2015). This work shows that greater variability in WIM over time is positively associated with mental illness, supporting the conceptualisation of WIM as an input, rather than a component, of mental illness. It is also supportive of the general idea that mental illness can be conceived as a higher order state/trait than WIM, which is consistent with dimensional affect models and differential emotions theory.

From a practical perspective, affirming the separability of WIM and anxiety speaks to the robustness of anxiety measures used in this study. If WIM affects mental illness reports this suggests that the constructs are not empirically distinct and are thus subject to measurement error (Rönkkö & Cho, 2020). Robust measures allow researchers and practitioners to make important distinctions between these facets and consider them as meaningful, independent of one another. Since the two constructs are significantly related to one another but do not share common variance, they should not be combined into composite measures, and it should not be assumed that measuring one can give you an accurate indication of the other. Importantly, the factors that may predict one may not necessarily predict the other and the two may combine in ways that intensify their effect on another construct (Payton, 2009). This kind of separability has been exemplified by research showing that having a mental disorder as well as poor mental health makes it much more likely that people will experience negative health consequences (Keyes, 2005). Given their separability, interventions should be separated into those targeting trait anxiety and those targeting WIM, and considered as distinct packages that work alongside, and contribute to one another. Research on state anxiety, given the disconnect between the present study and previous work, still needs to develop before firm conclusions can be drawn.

With respect to the impact of WIM on trait and state depression, however, an identified impact was observed in the present work. First, whilst discrete WIM conditions had no significant impact on trait depression post FDR corrections across both studies, increases in arousal change (in general across groups) did predict decreases in trait depression scores in study 2. This effect was small (Cohen's d = 0.106) but significant and in line with Associative Network Theory and H3, "**Low arousal**, negative valence (sad) **will predict greater increases in depression than anxiety relative to high arousal**, negative valence (anxiety)". Namely, given that depression is characterised by low arousal it follows that high arousal WIM will reduce scores on this measure, since low arousal cognitive associations will likely be harder to recall under these incongruent conditions. Interestingly, this impact was not identified in study 1, and counter expectations – "H1: Positive valence (happy and relaxed) will predict decreases in anxiety/depression relative to more negative valence (sad, anxious and neutral moods" - no impact was found in either study for positive/negative valence on trait depression. This suggests that WIM has a somewhat loose, and variable, overlap with trait depression that is determined by arousal but not valence. One implication of this could be that interventions targeting arousal (e.g., practicing relaxation) are more successful at reducing trait depression than interventions targeting valence (e.g., practicing gratitude). Future intervention research may benefit from exploring this possibility.

Second, whilst discrete WIM conditions also had no significant impact on state depression post FDR corrections, increases in valence change (in general across groups) did predict decreases in state depression scores in study 2 where state depression was explored. This effect was small (Cohen's *d* = 0.119) but significant and again in line with Associative Network Theory and H1, "**Positive valence** (happy and relaxed) **will predict decreases in anxiety/depression relative to more negative valence** (sad, anxious and neutral moods)". Namely, given that depression is characterised by negative valence, it follows that high valence WIM will reduce scores on this measure since negatively valenced cognitive associations will likely be harder to recall under incongruent affective conditions. This result is also consistent with previous work finding a significant impact in a negative direction for happy and a positive direction for sad mood (Baker, 1993).

It is interesting that valence and not arousal influenced state depression, and vice versa for trait depression. Whilst it is important for future research to check whether, and how often, these findings replicate, it could be the case that WIM has a different relationship with state and trait depression. If true, it may follow that interventions targeting positive valence will be more effective at relieving momentary depression and interventions targeting arousal will be more effective at relieving ongoing symptoms. This is an interesting premise for future research to test.

The observed conflation of WIM and trait/state depression also raises some important considerations. Note that measurement error is assumed to occur when two constructs are not empirically distinct (Rönkkö & Cho, 2020). Empirical distinctness between measures allows researchers and practitioners to make important distinctions between these facets and consider them as meaningful, independent of one another. Without this our theoretical understanding of how these two constructs arise may be misguided since the factors that predict one may also predict the other (Payton, 2009). A blurring of the lines between mood and mental illness may also result in trivialisation of mental illness which ought to be a distinct condition worthy of medical attention. Conflation of this kind can ignite concerns that ordinary emotional distress is being increasingly medicalised (Frances & Nardo, 2013) and contribute to the significant stigma surrounding mental illness (Mannarini & Rossi, 2019).

Whilst the effects observed in this study were small, they are not non-existent, and have highlighted the potential for WIM to impact upon depression reports. Since the effects of online mood inductions are reportedly weaker than those conducted in controlled lab settings, and since clinical populations have the propensity to experience mood much more intensely as well as experience more mood fluctuations in general, it is important that researchers pay attention to this result. Whilst it is beyond the scope of the present study to recommend the extent of empirical overlap that ought to be accepted between these constructs, highlighting the degree of overlap in existence currently can inform conversations and decision-making to this end. In the meantime, practitioners are advised to administer depression questionnaires more than once and aggregate results to avoid current mood-related bias in responses. This may be particularly important before providing any kind of diagnosis.

Study limitations

There are some notable limitations of these findings. First, since some of the WIM manipulations did not change arousal in the expected direction, the interpretation of results pertaining to the impact of arousal on mental illness reports is not straight forward. Namely, anxious mood is typically understood as being a highly aroused mood. Given this, it should have induced an increase in arousal, however this was not the case. It is important, therefore, that this work is replicated using other WIM induction methods (particularly those that may be more successful at raising arousal levels).

Second, since it was not possible to prove an implicit impact of WIM induction on positive and negative word evaluations, the possibility that reported changes in WIM were a function of demand effects cannot be eliminated. However, the fact that social desirability scores did not predict susceptibility to valence and arousal change does suggest this possibility is unlikely. It is perhaps more likely that the WIM inductions conducted were effective, but not strong enough to influence implicit evaluations of unrelated stimuli. Combined WIM induction procedures may be useful for future research to test here since these have been shown to have greater impact than other WIM induction tasks for aspects related to mental illness, such as dysfunctional cognitions (Van der Does, 2002).

Third, the present studies focused on the manipulation of WIM at just one time point and on two mental illness scales. These findings do not refute the possibility that recurrent WIM changes experienced over time may have a greater impact upon mental illness. Future work exploring how changes in WIM affect mental illness reports over time may also help to reveal a common threshold at which repetitive exposure to negative WIM experiences translates into mental illness. In a recent example of this prospect, research exploring the impact of 20-30 minute exposure to negatively valenced web content over a period of 5 days has shown a significant positive impact of negative valence on psychopathology, including "Anxious-Depression" (Kelly & Sharot, 2023). Finally, whilst a substantial proportion of people across both studies were categorised as having either mild, moderate or severe anxiety (62-66% across studies) and depression (31-40% across studies), future work can benefit from testing directly on clinical samples.

Conclusions

Together these studies have revealed that trait and state anxiety and depression measures are quite robust to WIM manipulations. However, there are some contexts where overlap between WIM and trait/state depression do exist. Researchers ought to consider this potential for overlap carefully when attempting to conceptualise and measure depression. Practitioners may also like to consider the implications of this overlap for depression intervention and diagnosis. Overall, this work presents a clear progression towards refinement in our ability to understand, conceptualise and measure two of the most psychologically debilitating conditions in existence.

Personal reflections on paper 2

These are personal research reflections and do not form part of the main paper. The purpose of this section is to highlight analysis challenges and key research skills developed.

Analysis challenges

One important analysis decision that had to be carefully considered was whether to use ANOVAs or regressions to assess the impact of the main result. In the end, given that they are considered superior for assessing the impact of categorical variables, I opted for ANOVAs. However, I also ran some more nuanced and complex analyses using the regressions with the raw valence and arousal change scores. What was particularly interesting here was that in some cases the results assessing the impact of discrete WIM-based variables on anxiety and depression differed for those assessing the impact of overall valence and arousal. Whilst this might simply have a been a result of more datapoints being available when assessing WIM across two continuous dimensions (vlanece/arousal) as opposed to two five categorical inputs (happy/sad/anxious/relaxed/neutral), it does raise some interesting questions about the relative utility of these approaches for understanding WIM. More specifically, it seems to support the view that dimensional assessments of WIM may be a more sensitive means of capturing and understanding the impact of WIM. Importantly, it may be that some important information is lost when confining WIM to discrete categories.

Another key challenge was deciding whether to opt for discrete condition-based WIM analysis or valence and arousal change analysis. The condition-based ANOVA was selected as the primary means of analysis for 2 reasons. First, and most importantly, it allowed for a comparison against the pre-specified neutral (control) condition which meant the impact of WIM inductions on mental illness could be assessed against a meaningful benchmark. Second, condition-based comparisons are the method that is typically selected in other WIM-induction studies which means the present results can be more easily compared to other research and may also reflect the best practice in the field. However, since I was interested in the comparison between these approaches, and I wanted to include additional factors of interest (e.g., personality and wellbeing) within the analysis for added robustness, I decided to run the regression analyses as a follow-up test. I believe that this added value and robustness to the results.

I analysed these data in R Studio. I trained myself using my LSE funded subscription to use Datacamp throughout this project so that I could address gaps in my coding knowledge. The gaps that I self-taught myself for this paper included how to set specific reference categories when running regressions, how to subset the sample to run sub-group analyses, and how to compute variable difference scores.

Key research skills developed

I ran these studies from conception to completion on my own. RCTs were chosen since it was necessary to isolate the impact of WIM induction on mental illness reports. Another option would have been to capture the impact of naturally occurring WIM on mental illness reports over time, however, an RCT is better suited to establishing causality and internal validity – defined as the ability to argue that observations are causal (Roe & Just, 2009) – in the first instance. A field-RCT can then be used as a follow-up to enhance the external validity, defined as the ability to generalise the relationships found in a study to other persons, times, and settings (Roe & Just, 2009), of any identified effects.

These experiments were initially designed to take place in the lab since WIM induction has been shown to be more effective under controlled conditions (Göritz, 2007). However, due to COVID-19 restrictions at the time the study was converted to an online experiment. Therefore, to address this concern participants were explicitly asked to complete the experiment from start to finish without breaks to reduce the influence of distractions.

One of the key things I learnt during this research was that accounting for people's scores along the valence and arousal dimensions of WIM can yield more information and predictive power than simply assessing the impact of discrete WIM conditions (e.g., happy/sad/relaxed/anxious). Future research assessing the impact of WIM on any given outcome can seek to benefit from this insight.

3.3. Paper 3

Pandemic related changes in social interaction are associated with changes in automatic approach-avoidance behaviour

Amanda Henwood*[1], Mike Rinck[2] and Dario Krpan[1]

*Corresponding author: Amanda Henwood (a.j.henwood@lse.ac.uk)
[1] Department of Psychological and Behavioural Science, London School of Economics and Political Science (LSE), United Kingdom
[2] Behavioural Science Institute, Radboud University, The Netherlands

Note: Methods come at the end of this paper due to formatting requirements of the journal to which it was submitted

Abstract

People's natural tendencies to either approach or avoid different stimuli in their environment are considered fundamental motivators of human behaviour. There is a wealth of research exploring how changes in approach and avoidance motivational orientations impact behaviour with consequences for wellbeing. However, research has seldom explored this relationship in reverse. The COVID-19 pandemic offered a unique opportunity to explore whether widespread changes in social behaviour produce changes in automatic approach-avoidance tendencies over time. We gathered online survey data on people's adherence to 7 of the prescribed social restrictions set out by the UK government and people's automatic approach-avoidance tendencies in response to different stimuli (sad/happy faces and social scenes) at three time points during the COVID-19 pandemic. Reduced-overall-interaction (digital and in person) was found to be significantly associated with faster avoidance relative to approach of sad faces. The results suggest that automatic approach-avoidance tendencies may function to protect people against the typically negative experience of reduced social

interaction, with important implications for understanding public resilience during times of crisis, and beyond.

1. Introduction

According to approach-avoidance theories, all behaviour can be conceptualised as a response to appetitive (rewarding) or aversive (punishing) stimuli (Elliot & Friedman, 2017). It is generally considered adaptive for humans to approach appetitive stimuli and avoid aversive stimuli. Changes in these approach-avoidance tendencies have been shown to characterise a host of behavioural and psychological disorders (Loijen et al., 2020). Identifying which behaviours and contexts facilitate adaptive and maladaptive approach-avoidance tendencies is therefore critical for understanding and promoting societal wellbeing.

Automatic approach-avoidance tendencies are typically captured using a joystick task (usually referred to as Approach-Avoidance Task, AAT) or a manikin task (called the Stimulus-Response Compatibility Task, SCRT – Loijen et al., 2020). In the former, individuals must either pull the joystick to enlarge images (approach) or push the joystick to shrink the images (avoidance) on the screen in front of them. In the SRCT, participants press a computer key to make a little manikin on the screen move towards (approach) or away from (avoid) the picture. Approach-avoidance tendency scores are computed by taking the difference between the time it takes people to avoid minus approach certain stimuli presented to them. The higher the tendency score, the faster people are to approach instead of avoid a given group of stimuli. For example, the so-called “sad tendency” is people’s tendency to approach sad stimuli faster than to avoid them.

Approach-avoidance tendencies are sensitive to various situational (e.g. threat), personal (e.g. anxiety) and behavioural (e.g. forced approach or avoidance of certain things in the environment) factors. For example, socially anxious people tend to evaluate neutral stimuli as more threatening relative to healthy individuals (Peschard & Philippot, 2017). As a result, they tend to be more avoidant of social stimuli, such as faces, than healthy individuals. Controlled approach-avoidance behaviour and subjective evaluations may differ from automatic (fast, efficient, goal-independent, and unconscious) approach-avoidance behaviour (Krieglmeyer et al., 2013). Socially anxious individuals, for instance, evaluate smiling faces just as positively as non-anxious controls. However, they show automatic avoidance of these smiling faces, being faster to avoid them than to approach them, whereas non-anxious individuals are faster to approach than to avoid them (Heuer et al., 2007). Similarly, abstinent alcohol-dependent patients avoid alcoholic beverages, and they report disliking alcohol, but nevertheless are faster to pull alcohol-related pictures closer than to push them away, unlike healthy participants (Wiers et al., 2011). This illustrates the need to study both controlled and automatic approach-avoidance behaviours, though separately from one another.

Using the joystick task or the manikin task, dysfunctional automatic approach-avoidance tendencies have also been found in cannabis and nicotine addiction, eating disorders, several anxiety disorders, and in depression (Loijen et al., 2020). The importance of these dysfunctional tendencies in psychopathology is illustrated by the fact that they can be modified by extensive joystick task trainings (a specific type of Cognitive Bias Modification, CBM), which then can lead to reduced psychopathology. In these trainings, participants usually use a joystick to complete many hundred trials of pulling closer pictures of stimuli they automatically avoid or fail to approach (as in anxiety disorders and depression, respectively). The opposite is done in addictions, where automatic drug-avoidance is trained by having patients complete many hundred trials of pushing away pictures of the relevant drug. For instance, a 6-session training during which pictures of alcoholic beverages are always pushed away with a joystick, and pictures of non-alcoholic beverages are always pulled closer, reduces relapse rates in currently abstinent alcohol-dependent individuals by about 10% (Loijen et al., 2020).

Similarly, training socially avoidant individuals to approach pictures of smiling faces (which they would typically avoid), has been shown to promote more adaptive behavioural responses (Taylor & Amir, 2012), Identification and alteration of these endogenous factors has led to significant progress in understanding and promoting individual wellbeing (Loijen et al., 2020).

However, all of this research addressed existing dysfunctional approach-avoidance tendencies without being able to identify their origin (when comparing patients to healthy controls), or the tendencies were modified directly via training (when CBM training protocols were tested), or tendencies were assessed under carefully controlled laboratory conditions (when responses to conditioned stimuli were measured). In contrast, nothing is known about how significant real-world influences might be associated with the approach-avoidance tendencies shown by large, unselected samples from the general population.

The COVID-19 pandemic provided a unique opportunity to test exactly this. Since its outbreak in 2019 COVID-19 has caused more than 6 million deaths worldwide (WHO, 2022). To limit the spread of the virus many governments encouraged, or in some cases enforced, considerable restrictions on social behaviours. Social distancing, self-isolation, mask-wearing in public, and reduced social interaction and gatherings were amongst the most popular of these measures. These policies facilitated a marked shift in social behaviour across the globe (Dryhurst et al., 2020). As such, the crisis offered the opportunity to explore the impact of widespread behaviour change on approach-avoidance tendencies. Utilising this, the present longitudinal study explored whether changes in social behaviour can be associated with changes in automatic approach-avoidance tendencies in a natural setting.

COVID-19 policy responses were determined in large part by a focus on avoiding mortality risks and deterioration of physical health (Hsiang et al., 2020; Qiu et al., 2020; BMJ, 2022). However, to truly understand the full impact of any behavioural measures put in place to alleviate physical health risks, it is necessary to obtain a clear vision of the impact of such measures on psychological health, too (Dolan et al., 2021). A growing number of studies have highlighted strong associations of psychological behaviours such as social distancing with mental health and wellbeing. A US study conducted in March 2020, as the pandemic grew worse and stay-at-home orders were issued, found social distancing behaviours to be associated with increases in anxiety and depression (Marroquín et al, 2020). In the UK, a large longitudinal study assessing the trajectories of anxiety and depression over the 20 weeks after lockdown was announced, found that anxiety and depression were particularly high at the start of these measures being introduced (Fancourt et al., 2021). However, these high levels of anxiety and depression declined rapidly thereafter, suggesting successful adaptation to the measures over time by many. Whilst circumstantial risk factors such as being female, young, and having lower educational attainment have been explored, the processes underpinning such impacts and adaptation remain underexamined.

Approach-avoidance tendencies are an important mechanism through which reductions in social behaviours might impact wellbeing. For example, people may respond to reduced social interaction by altering their approach-avoidance tendencies in response to valanced stimuli, which have been noted as playing an important role in emotion regulation. For example, people may increase their approach of positive stimuli to buffer against the negative impact of that lost interaction. These changes in tendencies and their psychological implications can help to shed light on important wellbeing consequences of enforced social distancing. Another possibility might be that people increase their approach relative to avoidance of social stimuli. In theory, such increased motivation towards social stimuli should increase the likelihood that people will encounter and engage with social opportunities, allowing them more opportunity to make up for lost social interactions. If social distancing measures lead to an increased tendency to socially interact, this may prevent social distancing measures from taking effect. This would result in a self-defeating policy and thus warrants our close investigation. Identifying patterns such as these and assessing their relationship with affective outcomes can help us to better understand which kinds of behavioural responses are conducive (or otherwise) to building psychological resilience in response to reduced social interaction. Since automatic tendencies can be trained over time, interventions that encourage individuals to approach social stimuli, could then be used to protect vulnerable individuals against psychological decline.

Therefore, this research is important not only to help us understand more about the interrelationship between behaviour and approach-avoidance tendencies in general, but also to help governments worldwide build a clear picture of the extent to which policies that encourage reductions in social behaviour, may impact the psychological health of their citizens. It can also help to identity pathways through which to mitigate the negative impacts of reduced social interaction, and loneliness, more generally.

The present work and theoretical framework

Only a few studies so far have investigated approach and avoidance in the context of the COVID-19 pandemic. In summary, these found that individual differences in approach, relative to avoidance, were more important predictors of compliance with COVID-19 recommendations, and were positively associated with social distancing, wearing masks and gloves, and reduced mobility (Krupić et al., 2021). Moreover, individual differences in avoidance were associated with impaired wellbeing during the pandemic (Shamblaw et al., 2021), and self-reported mask-related worrying (e.g. feeling nervous when seeing people in masks) was associated with lower avoidance bias toward unmasked people, but only for participants with low COVID-19 anxiety (Krishna et al., 2021). However, our study is the first to investigate whether changes in isolated, in addition to more general, social behaviours are associated with changes in approach-avoidance tendencies.

To examine this, we measured adherence to some of the main restrictions on social behaviours set out by the UK government (see Table 1 below) as well as approach-avoidance tendencies in response to sad/happy faces and social scenes at three different time points over three months. We gathered these data during a period of lockdown easing in the UK (May – July 2020) following the country's first strict lockdown in March (see Methods Table 7 for more detail of lockdown context across waves). All data were gathered using online surveys and approach-avoidance was measured using the Stimulus Response Compatibility Task, which required participants to press a computer key to make a manikin on the screen move towards (approach) or away from (avoid) the picture.

Table 1. Social distancing behaviours measured

1. Social distancing item	I have been social-distancing (keeping 2 metres apart from other people outside of the house)
2. Self-isolating item	I have been self-isolating (i.e. not being in contact with anyone else or leaving the house)
3. Avoiding crowds item	I have been avoiding crowds
4. Avoiding small groups item	I have been avoiding small group face-to-face activities with friends and family from outside of my house
5. Reduced in-person interaction item	I am having less in-person social interaction than before social distancing measures were first introduced
6. Reduced-overall-interaction item	I am having less overall social interaction (digital or in person) than before social distancing measures were introduced
7. Mask outdoors item	I have been wearing a mask outdoors

There are two main theoretical accounts that can inform our predictions: compatibility hypothesis (Neumann & Förster, 2014) and emotion regulation theory (McRae & Gross, 2020). According to the compatibility hypothesis that has appeared in various formulations across many articles (Alexopoulos & Ric, 2007; Huntsinger, 2013; Krieglmeyer et al., 2013; Seibt et al., 2008; Schmitz & Wentura, 2012; Strack & Deutsch, 2004) negative (positive) affect, emotions, and experiences are compatible with avoidance (approach) motivation and should thus facilitate avoidance (approach) responses. Therefore, avoidance (approach) can be conceptualized as a preparedness to respond to negative (positive) objects (Strack & Deutsch, 2004). For example, people are faster to push rather than pull various negatively valenced stimuli, from negative words (Chen & Bargh, 1999) to images of spiders they are afraid of (Rinck & Becker, 2007). These positive/negative stimulus evaluations are usually grounded in innate or learned tendencies (Monni et al., 2020). In that regard, we might expect a link between reduced social behaviours and reactions to different faces that complies with an extended version of the compatibility hypothesis. For example, reduced social interaction may be an inherently negative experience given its adverse consequences for mental health (Loades et al., 2020; Pancani et al., 2021; Loijen et al., 2020; Ganesan et al., 2021) and thus further potentiate avoidance responses to the compatible negatively valenced faces. In other words, if reduced social behaviours are negatively valenced experiences, they may simply predispose people to react more quickly to avoid sad faces as they are emotionally congruent stimuli.

Emotion regulation theory broadly refers to people's attempts to influence emotions in themselves and others (Gross, 2015; McRae & Gross, 2020; Thompson, 2011). There are five families of emotion regulation strategies, and two of them are particularly relevant in the context of the present research: situation modification and attentional deployment (McRae & Gross, 2020). Situation modification involves undertaking action to change a given situation to experience desired emotions, whereas attentional deployment involves directing attention away from (or toward) stimuli that evoke undesirable (desirable) emotions. Therefore, it is plausible that people may undertake approach-avoidance reactions regarding images of faces, and this could fall both under situation modification (i.e., changing the situation by pushing or pulling specific faces) and attentional deployment (i.e., pulling certain faces toward oneself or pushing them away to direct attention toward or away from these faces).

In that regard, an argument could be made that, if people naturally associate reduced social behaviours with feeling negative, they may experience an inclination to approach (vs. avoid) happy faces and avoid (vs. approach) sad faces. The idea being that this inclination functions to protect them against the potential for experiencing negative feelings in the future. Therefore, the present work focused on testing for a direct association between changes in social behaviours with changes in approach-avoidance responses. Individuals may also have stronger inclinations to approach (vs.

avoid) social stimuli more generally, given that they may perceive assuaging the need for social interaction by exposing themselves to social stimuli as a response that will bring about positive feelings. Similar self-regulating feedback loops between approach-avoidance tendencies and affect have been proposed by prominent approach-avoidance theorists and thus warrant testing in a natural setting (Carver, 2006).

Overall, both the compatibility hypothesis (Neumann et al., 2003) and emotion regulation theory (McRae & Gross, 2020) imply that approach-avoidance tendencies can be meaningfully shaped by negative experiences. Since many studies have shown that reducing social behaviour is an inherently negative experience, we predict that reduced social behaviour will shape approach-avoidance tendences toward different social stimuli. However, the two theories yield different predictions. As discussed, the emotion regulation theory predicts that reduced social behaviours will be linked to both stronger approach relative to avoidance tendencies toward happy faces (**H1**), and to stronger avoidance relative to approach tendencies regarding sad faces (**H2**), as a strategy to maintain a more optimal emotional state. Similarly, the theory predicts that reduced social behaviours will be linked to stronger approach relative to avoidance tendencies toward social stimuli (**H3**), given that these behavioural reactions may serve to activate desired motivational states by assuaging the need for social interaction. In contrast, the compatibility hypothesis predicts only **H2**, given the compatibility of negatively valenced sad faces with avoidance tendencies and reduced social behaviours that are inherently negative.

Analysis plan

The aim of the analysis was to explore whether changes in social behaviours are significantly associated with changes in approach and avoidance behaviour over time. Since the nature of this study was exploratory, we began by first checking descriptive statistics for all input and outcome variables. We then ran three balanced, fixed effects panel linear models on the respective approach-avoidance outcomes of interest (happy, sad, and social tendency) using the 7 prescribed social behaviours as predictors to assess whether any significant relationships were present. Panel linear regression was chosen since it is free from distributional assumptions (PLM, 2022). Each of the four tendencies was computed as the difference between the participant's mean response time to start a correct manikin movement away from the corresponding pictures (avoidance) minus the mean response time to start a correct movement towards the pictures (approach). For instance, the sad tendency reported below was computed as: Mean reaction time (RT) to start a movement away from sad faces minus mean RT to start a movement towards sad faces. Positive values of these tendency scores reflect relative approach of the stimulus category, whereas negative values reflect relative avoidance. In previous studies, typical ranges of these scores were between -50 ms and +50 ms.

To account for multiple testing, we applied the False Discovery Rate Controlling Procedure to all reported p-values below (Glickman et al., 2014). Additionally, to account for possible multicollinearity issues arising from correlations between the social behaviours (even though these correlations were not decidedly large (Yoo et al., 2014), Appendix Table 1), we ran separate models for each of the 3 outcomes using each single behaviour as a predictor on its own to determine whether this changed any results from significant to insignificant or vice versa. To test whether affective variables – or any other of the measured time-varying variables – were partially responsible for any observed effect, we also re-ran the singular behavioural predictor model with all additional time-varying covariates measured in the study.

The covariate model included several key affective variables, including anxious, happy, and stressed today, valence, arousal, social anxiety, anxiety, and life satisfaction. These variables were included to control for their potential influence on the relationship between reduced overall interaction and sad tendency. For example, reduced-overall-interaction might only affect sad tendency because of its impact on negative affectivity. If this were the case, then removing variables associated with negative affectivity should alter the strength of the association between reduced-overall-interaction and sad tendency. Thus, including affective variables helps us to partition out the unique variance associated with these variables.

Fear of contracting COVID-19 both for oneself and for others were included to account for the potential impact of the COVID-19 context on the relationship between reduced overall interaction and these tendencies. For example, it might be that reduced overall interaction is highly associated with fear of contracting COVID-19, and it is this fear rather than the reduced interaction itself that impacts upon sad tendency. In addition, Behavioural Inhibition System (BIS – corresponds to motivation to avoid aversive outcomes) and Behavioural Approach System (BAS – corresponds to motivation to approach goal-orientated outcomes)) tendencies were included to control for people's general sensitivity towards positive and negative stimuli (Carver & White, 1994) and especially since they have been previously associated with infection avoidance (Bacon et al., 2022).

Since there was no obvious reason for us to favour a fixed over a random effects model based on our data structure, we ran both covariate models and conducted the Hausman test to compare model fit (Amini et al., 2021). We also ran a random effects plm with non-time varying variables (age, gender, ethnicity, income, region, mental health, physical health, subjective wellbeing, Big 5 personality traits, number of children, number of people living in the house, keyworker status and COVID-19 symptoms) since inclusion of time non-varying variables was not possible with fixed effects plms. This helped to account for the possible impact of individual differences on any observed association between social behaviour and approach-avoidance and partial out any variance associated with these more general factors.

To improve generalisability and understanding of these data, the covariate models listed above were conducted on three dataset variations: 1) balanced panel dataset including only participants with data at all three time points, 2) maximum sample dataset that included participants with data from 2 time points or 3 time points, 3) maximum change dataset that included only participants with data from time points 1 and 3 , where the largest changes in social behaviour had taken place.

To account for the possibility that it is not the social behaviours themselves but rather participants conscious feelings (e.g. conscious feelings regarding their day yesterday, how they feel right now, or more generally) that may be associated with these behaviours driving any observed effect, we also tested whether negative affective variables were significantly associated with social behaviours in separate fixed effects panel linear models, for instances where a significant association between social behaviour and approach-avoidance was found. Finally, to obtain a more complete understanding of the interrelationships between social behaviours, negative affectivity, and approach-avoidance, we tested for an association between our measured affective variables with the relevant approach-avoidance tendency, also using panel linear models (e.g., anxious today predicting sad tendency).

2. Results

Descriptive statistics

Below we report the descriptive statistics for sad tendency and social behaviour 6: reduced overall (digital and in person) interaction. We report these variables since, as can be seen in section 3.2. below, a statistically significant relation was found between them, and they are thus the central focus of this paper. Descriptive statistics and further analyses of the other tendency outcomes are reported in the appendix (Tables A. 2-8).

Table 2. Sad tendency means across waves

Wave	Mean	SD	Count
1	6.102	324.141	174
2	-36.771	302.293	174
3	2.292	285.294	174

As can be seen in Table 2, people became more avoiding of sad images on average in wave 2 as compared to waves 1 and 3. However, a repeated measures ANOVA with time as the independent variable and sad tendency as the dependent variable indicated that these changes were not significant ($F(2,346) = 1.161$, $p = 0.314$).

Table 3. Reduced-overall-interaction means across waves

Wave	Mean	SD	Count
1	7.023	3.149	174
2	6.665	3.151	174
3	6.483	3.147	174

As per Table 3, people reported lower values for reduced-overall-interaction on average over time, meaning that as time went on and lockdown restrictions eased, people interacted more. However, a repeated measures ANOVA with time as the independent variable and reduced-overall-interaction *overall* as the dependent variable indicated that these changes were not significant $F(2,346) = 1.782$, $p = 0.17$).

Fixed effects panel linear model results

In the fixed effects panel linear model exploring the impact of the 7 behavioural predictors on sad tendency, we found that reduced-overall-interaction was a significant predictor of sad tendency (tendency to approach over avoid sad faces, see Table 4). Specifically, it was found that reduced overall interaction was significantly associated with stronger avoidance relative to approach of sad faces over time. With each one-unit change in reduced-overall-interaction (measured on a 0-11 scale), sad tendency reduced by about 18 milliseconds. This relation remained significant following FDR correction. It also remained significant when ran as fixed effects model with less overall interaction included as the only predictor variable (Table A.9), and when ran as a random effects panel linear model (Table A.10).

Table 4. Simple fixed effects panel linear model showing a significant association between reduced-overall-interaction and sad tendency

	Estimate	Std. Error	P
Social distancing	7.619	12.726	0.550
Self-isolating	-8.906	6.316	0.159
Avoiding crowds	12.413	13.317	0.352
Avoiding small groups	-2.674	7.562	0.724
Reduced *in person* interaction	-0.226	8.529	0.979
Reduced *overall* interaction	-17.811	6.426	0.006**
Mask outdoors	2.645	6.057	0.663
N	174		
F	1.922		
R^2	0.038		
Cohen's *d* for reduced overall interaction	0.210		

Note: Fixed effects regression using prescribed social behaviours (independent variables) to predict sad tendency (dependent variable). Standard errors are clustered on an individual level. Cohen's *d* value was calculated using the reduced-overall-interaction coefficient to estimate effect size. * p<0.05, ** p<0.01, *** p<0.001

Overall, this result was consistent with the predictions made from the compatibility hypothesis (**H2**: reduced social behaviours will be linked to stronger avoidance relative to approach tendencies regarding sad, but not happy or social, faces) given the affective compatibility.

In all three fixed effects covariate panel linear models, the significant association between reduced-overall-interaction and sad tendency remained (Table 5). The strongest model was model 3 which was conducted on the maximum change dataset (R^2 = .139). In this model, a one unit increase in reduced-overall-interaction (0-11 scale) was associated with a 26.555 millisecond decrease in sad tendency (tendency to approach over avoid sad faces). Therefore, reduced overall interaction was associated with higher avoidance relative to approach of sad stimuli.

Table 5. Reduced-overall-interaction is significantly associated with sad tendency when all time-varying variables are included in the fixed effects panel linear model

	Dependent variable:		
		Sad tendency	
	Balanced	Maximum sample	Maximum change
Reduced-overall-interaction	-18.733***	-17.281***	-26.555***
	(5.713)	(5.483)	(7.592)
Corona fear	10.922	10.193	-11.547
	(15.246)	(13.993)	(18.601)
Corona fear others	-8.934	-7.989	9.447
	(14.202)	(12.862)	(18.129)
Anxious today	4.063	5.689	25.528
	(12.147)	(11.416)	(16.431)
Happy today	-25.855*	-26.631*	-2.842
	(15.491)	(14.857)	(22.782)
BIS	4.940	2.284	8.617
	(7.190)	(6.963)	(8.773)
BAS Drive	6.650	6.279	4.002
	(10.061)	(9.482)	(12.493)
BAS Reward	5.582	4.856	1.384
	(5.894)	(5.486)	(6.513)
Hours away from home	-0.057	-1.974	14.447
	(15.622)	(14.847)	(20.236)
Stressed today	-25.596**	-28.009**	-39.730**
	(12.246)	(11.575)	(15.626)
Social Anxiety	-0.536	-0.887	-1.267
	(1.019)	(0.967)	(1.342)
Anxiety	-8.998**	-4.926	-11.900**
	(3.960)	(3.684)	(5.486)
Life Satisfaction	11.111	13.994	17.739
	(21.277)	(20.394)	(29.361)
Valence	0.530	0.206	-0.956
	(1.179)	(1.069)	(1.466)
Arousal	-1.265	-1.098	-1.040

	(0.896)	(0.851)	(1.188)
N	522	608	415
R^2	0.075	0.059	0.139
Adjusted R^2	-0.456	-0.548	-1.015
F Statistic	1.784** (df = 15; 331)	1.542* (df = 15; 369)	1.899** (df = 15; 177)
Cohen's *d* for less overall interaction	-0.144	-0.138	-0.153

Note: *p<0.1; **p<0.05; ***p<0.01

Fixed effects regression using prescribed social behaviours (independent variables) and other time varying covariates to predict sad tendency (dependent variable). Standard errors in brackets are clustered on an individual level. Cohen's *d* value was caculated using the reduced-overall-interaction coefficient to estimate effect size.

Random effects panel linear model results

In the random effects covariate panel linear model including additional time-varying predictors (Table 6), the significant impact of reduced-overall-interaction on sad tendency also remained. This significant association also remained when including time non-varying variables into the random effects model (Table A.11).

Table 6. Reduced-overall-interaction is significantly associated with sad tendency when all time-varying variables are included in the random effects panel linear model

	Dependent variable:		
	Sad tendency		
	(1)	(2)	(3)
Reduced-overall-interaction	-15.599***	-12.517***	-15.828***
	(4.509)	(4.216)	(5.095)
Corona fear (self)	1.872	1.915	-1.850
	(7.830)	(6.961)	(8.544)
Corona fear (others)	-1.298	-1.657	2.621
	(8.437)	(7.419)	(9.235)
BIS	-4.577	-5.955	-7.168
	(4.660)	(4.243)	(4.975)
BAS Drive	7.539	4.023	-2.283
	(6.351)	(5.847)	(7.074)

BAS Reward	-1.160	0.268	-0.671
	(3.330)	(3.152)	(3.465)
Stressed today	-10.190	-11.451	-15.544
	(9.353)	(8.709)	(10.386)
Anxious today	12.143	11.047	18.442*
	(9.039)	(8.341)	(10.206)
Happy today	-14.824	-15.896	-12.072
	(11.968)	(11.210)	(14.406)
Social Anxiety	-0.054	-0.271	-0.585
	(0.563)	(0.518)	(0.636)
Anxiety	-0.146	1.713	1.069
	(2.278)	(2.080)	(2.537)
Life Satisfaction	6.249	10.418	16.028
	(14.359)	(13.149)	(16.502)
Valence	0.972	0.692	-0.401
	(0.904)	(0.816)	(0.984)
Arousal	-0.970	-0.720	-0.304
	(0.615)	(0.567)	(0.710)
Hours away from home	14.044	11.595	12.302
	(11.732)	(10.747)	(12.971)
Constant	144.826	106.980	227.751
	(194.515)	(179.019)	(216.393)
Observations	522	598	415
R^2	0.040	0.028	0.045
Adjusted R^2	0.011	0.003	0.010
F Statistic	20.816	16.999	18.974
Cohen's d for reduced-overall-interaction	-0.151	-0.130	-0.134
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: Random effects panel linear regression using reduced-overall-interaction and all other time varying covariates (independent variables) to predict sad tendency (dependent variable). Standard errors for each estimate are listed in brackets. Cohen's *d* value was caculated using the reduced-overall-interaction coefficient to estimate effect size.

The Hausman test revealed that neither model produced significantly different results (p = 0.163), therefore the random effects model was chosen as the best specification for our data.

Relationships between affective variables with reduced-overall-interaction and sad tendency

Counter to expectations, none of the affective variables were significantly associated with reduced-overall-interaction in our fixed effects panel linear models (Tables A.12-14).

In the fixed effects panel linear models exploring whether changes in affective variables were associated with changes in sad tendency, we did find some significant effects (Tables A.15-17). Increases in generalised anxiety levels were associated with less approach over avoidance of sad faces (β = -8.703**). Therefore, the more anxiety people experienced over time, the faster they were to avoid (relative to approach) sad faces. We also found a significant negative association between stressed today and approach over avoidance of sad faces (sad tendency - β = -19.030**). Therefore, higher levels of stress were associated with less approach (relative to avoidance) of sad images. We note that these significant associations were not present for happy for social tendency. Neither fear of infecting oneself nor of infecting others with the corona virus was associated with sad tendency.

3. Discussion

COVID-19 facilitated an unprecedented shift in social behaviour across the globe. This study used the prescribed changes in behaviour set out by the UK COVID-19 policy guidelines as a natural setting to assess whether changes in social behaviours are associated with changes in approach-avoidance tendencies over time. To this end, during a period of lockdown and easing of social restrictions, we asked people about the extent to which they had been following several socially relevant policy guidelines over the last 7 days. We also measured their approach-avoidance tendencies in response to sad/happy faces, and social scenes, at three time points, each one month apart. Research of this kind can provide the best possible evidence for relevant real-world phenomena when random assignment of participants to experimental conditions is not possible. We found a significant relationship between “less overall (digital and in person) social interaction since social distancing measures were first introduced” and participants’ tendency to approach over avoid sad faces (sad tendency). Specifically, reducing overall social interaction was related to more avoidance relative to approach of sad faces. The other prescribed social behaviours (social distancing, self-isolating, avoiding crowds, avoiding small groups, less in person interaction and wearing a mask outdoors) had no significant relationship with approach-avoidance tendencies.

Up until now, studies demonstrating the impact of real-world behaviour changes on approach-avoidance tendencies have been lacking. The relationship between behaviour and approach-avoidance has typically been shown in the inverse direction, whereby training people to alter their automatic approach-avoidance tendencies succeeds in altering their real-world behaviour (Krypotos et al., 2015; Sharbanee et al., 2014; Asnaani et al., 2014). This lack of research is likely because implementing and capturing widespread behaviour change over time can be challenging. To our knowledge, this research is the first of Its kind to document a relationship between widespread behaviour and automatic approach-avoidance change in a natural setting. According to our most effective model (R^2 = 0.139), with each one-unit change in reduced-overall-social-interaction (measured on a 0-11 scale), sad tendency reduced by about 26 milliseconds. Whilst this effect size is small-medium (Cohen’s d = 0.153) the magnitude of this change is generally understood to be meaningful in the approach-avoidance literature ((Funder & Ozer, 2019; Loijen et al., 2020). Moreover, the magnitude is similar to the one frequently achieved by direct attempts to modify approach-avoidance tendencies, suggesting that continued investigations of the relation between real-world behaviour change and changes in automatic approach-avoidance tendencies are worthwhile.

Importantly, the negative association between reduced overall interaction and sad tendency remained robust to three different dataset variations as well as different model calculation methods (fixed and random effects). It also survived FDR correction; therefore, we believe the identified effect is a robust one. Reduced overall interaction was most strongly associated with sad-face avoidance in the models that included other time-varying predictors, relative to the models that only included the 7 behavioural predictors. This is likely because many of the prescribed social behaviours captured will have shared some variance with one another. Therefore, we considered the covariate model results to be a more accurate and robust representation of impact. However, the model including all 7 prescribed social behaviours was important to determine whether any of the behaviours in question were significantly associated with approach-avoidance tendencies over and above the other behaviours, which was the case for reduced-overall-interaction.

We consider the possible impact of reduced overall interaction on sad-face avoidance (and not happy face or social scene avoidance) to be consistent with the compatibility hypothesis insofar as our Hypothesis 2 was supported. If people were adapting their automatic approach-avoidance tendencies to regulate their negative affectivity in response to reduced social behaviours as per emotion regulation theory, then we would have expected to see an association between reduced social behaviours with happy and social, in addition to sad, tendencies. We would also have expected to see an association between reduced-overall-interaction and the affective variables, which we did not. Instead, our results appear to be a better fit with the compatibility hypothesis, which contends that people respond to affective experiences with compatible approach-avoidance responses. For example, somebody feeling negative should be faster to respond with avoidance to negative images. We specifically predicted therefore, that since reduced social interaction is an inherently negative experience for humans to undertake, it would facilitate faster compatible responses to negative images (e.g., faster avoidance relative to approach of sad faces) (H2).

Since reduced-overall-interaction was not associated with any of the affective self-report measures, it is possible that the link between reduced-overall-interaction and increased avoidance relative to approach is direct and cannot be explained by conscious affective experiences such as WIM. It may be the case then that our automatic approach-avoidance based response system causes an immediate response which serves to protect us against downstream negative effects of extreme isolation before they have taken hold. Our research is therefore aligned with the prospect that automatic approach-avoidance responses function as a precursor to affective responses (in addition to being driven by them).

Consistent with more general findings, we did find an important association between changes in affect and changes in approach-avoidance over time. However, to our knowledge, this is the first time that such associations have been demonstrated in a natural setting. Namely, we found significant negative associations between changes in generalised anxiety over time, and changes in momentary stress levels over time, with sad tendency. That is, increases in general anxiety and stress yesterday were associated with decreases in people's tendency to approach over avoid sad images. This latter finding is also consistent with the compatibility hypothesis since we observe an association between negative affective traits/states and compatible avoidance responses to negative stimuli (sad faces) and not positive or social stimuli (happy faces and social scenes). This avoidance of negative stimuli in response to negative affective states/traits may function in a similar way to the processes observed in optimism bias research. This research finds that healthy individuals show a bias towards more desirable stimuli which is thought to enable healthy psychological functioning (Sharot, 2012). Since faster avoidance relative to approach of aversive stimuli is generally considered adaptive, this tendency might have again increased to bolster individuals against downstream reductions in mental wellbeing (Matias et al., 2020).

Importantly, reductions in in-person interaction alone were not significantly associated with sad tendency, rather it was only the combination of both digital and in person interactions (reduced-overall-interaction) that had an impact. Presumably, circumstances where people substitute in person for digital interactions may negate the need for a mechanism to reduce any negative affect that arises from that loss of interaction. Indeed, research has shown that digital interaction can buffer against the negative psychological impacts of social isolation (Sen et al., 2022), perhaps due to its ability to function as an alternative social outlet (Grieve et al., 2013). This highlights the potential importance of online platforms that enable communication during periods of social restrictions. It also suggests that those unable to shift their social interactions to an online format may be at increased risk from policies that restrict face-to-face interactions and thus ought to be prioritised by government interventions seeking to mitigate the negative impact of the pandemic on the public. This substitution of reduced in person social interaction with online interactions can likely also explain the null impact of the other social behaviours on approach-avoidance tendencies. We recommend that future work seeks to disentangle the effects of online compensation for reduced social interaction in an experimental setting.

There are some notable limitations to this study. First, this study cannot prove a causal link between reduced-overall-interaction and sad tendency since behavioural and approach-avoidance variables were collected at the same three time points rather than one after the other. Whilst panel linear regressions are proficient at controlling for time invariant unobserved variables (Jin, 2017), they cannot control for time varying omitted variables. For example, one possibility is that people who obeyed the rules more had a less positive outlook on COVID-19 and therefore more sad tendency. Though, the potential for this bias is less in panel data than it is in cross-sectional data. Nevertheless, it is possible that some unobserved variables or external events might explain the association. Second, in the present study, we focused on reducing type I error (i.e., false positive findings) by implementing FDR corrections (Benjamini & Hochberg, 1995), given that this error is responsible for the replication crisis in psychology (Oberauer & Lewandowsky, 2019). However, we acknowledge that our emphasis on type I error may have inflated type II error (Shrout, 2012). In other words, we may have potentially failed to detect certain "real" effects because the corrections we used raised the bar for detecting an effect. Therefore, the stringent approach that we used may have prevented us from detecting certain smaller effects that could have potentially provided more nuanced insights into the link between reduced social behaviours and approach-avoidance tendencies.

Moreover, since we did not see significant changes in reduced-overall-interaction over time, it is likely that the effect we observed may be underestimated, and therefore greater under conditions where significant changes are present. This is consistent with our finding that the association between reduced-overall-interaction and sad tendency becomes stronger when considering responses that were further apart in time. In terms of generalisability, it is possible that self-reported behaviours may not correspond perfectly to actual behaviours. Additionally, though we ran many tests to determine the robustness of the identified association between changes in social interaction with changes in sad tendency, it is important to note that most of the other social behaviour associations did not reach statistical significance. Thus, it is strongly recommended that the identified association is retested in future work. For example, lab-based studies may be used to test the impact of experimentally manipulated social isolation (see Baumeister et al., 2002 and Cacioppo et al., 2006 for examples) to better establish causality and ensure that this association can be generalised to non-COVID-19-related contexts.

It must also be noted that the necessity for a rapid response to an evolving situation meant it was not possible to conduct pre-validation studies on the social scenes used in this study. Therefore, future work should also seek to validate the social scenes presented since insufficient stimuli selection might have contributed to null results for these stimuli. Finally, it would have been preferable to test for the affective variables as mediators. However, there are some major flaws associated with mediation in the case of our research design and so we did not include this in our analysis[1,2]. Namely, since we tested behaviours, approach/avoidance tendencies, and affective variables at the same time points, we can't say with certainty what precedes what. Due to the issues with mediation, we included important variables as covariates and we additionally explored the association between the variables we highlight as being potentially important for the relationship between social behaviours and approach-avoidance tendencies.

Overall, this study has revealed a potentially important relationship between deliberative avoidance of social behaviours and automatic approach-avoidance tendencies. Namely, during the COVID-19 pandemic people appear to have responded to reduced social interactions by increasing the extent to which they avoided (relative to approached) sad stimuli. This underscores the potential adaptive significance of approach-avoidance tendencies in response to behaviour, rather than just as determinants of behaviour. Existing research on differences in approach-avoidance tendencies has focused either on stable differences caused by stimulus types (e.g., pleasant vs. unpleasant stimuli such as spiders vs. butterflies) or on stable inter-individual differences (e.g., spider fearful vs. non-fearful, patients vs. healthy controls). The current study shows that is also worthwhile to investigate temporal changes and intra-individual differences.

It is highly possible that such findings, although observed during the COVID-19 pandemic, may also refer to isolation and reduced social interaction more broadly. Policymakers may use these data as a starting point to better understand the potential impacts of social isolation, and identify individuals who may be at increased risk of declining wellbeing in response (e.g., those who are unable to compensate in person interaction with digital interaction). Though, we stress the importance of targeting social isolation at the source by increasing social activity wherever possible whilst retaining these techniques for contexts where increasing interaction is not an option (e.g., during national lockdowns or for incapacitated individuals).

Researchers can use these data as a foundation from which to design new experiments and interventions that test the causality of these associations, for example, by seeing whether training avoidance of sad stimuli can help to increase wellbeing in the face of loneliness. If successful, policymakers can administer such trainings to members of the public at high risk of social isolation. As such, we encourage researchers interested in behaviour, wellbeing, and motivation, to embark on a new line of research that further explores the role of approach-avoidance tendencies as adaptive responses to changes in wellbeing-related behaviours.

4. Methods

Data collection

This study ran between 23rd May and 20th August 2020. We sampled from a nationally representative online population in the UK at three time points over three months. Whilst the study was UK based, the social behaviours investigated were consistent with those put in place by many other countries across the globe. A month between waves was selected as the time period over which changes in social behaviour were expected to have taken place due to the easing of lockdown restrictions. The dates and lockdown contexts for each of the three waves is detailed in Table 7 below. Participants were recruited via PureProfile online study recruitment agency and received £4 for participating in the study. All methods were performed in accordance with the relevant guidelines and regulations. The experimental protocol was approved by the London School of Economics Ethical Committee (ref#1133). Informed consent was obtained from all participants.

Table 7. Dates of each wave and associated UK COVID-19 restriction context (Greater London Authority, 2021)

	Date	Lockdown status
Wave 1	23rd May 2020	First lockdown has started to ease, people are allowed to leave the house to sunbathe and exercise more than once a day. People must keep two metres away from others and are also encouraged to wear face coverings in enclosed places. It is not possible to meet others in groups, any schools are closed, and non-essential shops are shut.
Wave 2	22nd June 2020	Virus alert level downgraded from four to three. Schools have gone back. Non-essential shops are back open. People are now allowed to meet outside in groups of up to 6.
Wave 3	21st July 2020	Work-from-home guidance eased as England plans for return to normality. Pubs are open again and weddings allowed. On 24th July face covering becomes mandatory

Participants

There were 1097, 325 and 267 consenting participants obtained in each wave. From this pool, we removed those who: 1) did not have both survey and approach avoidance data, 2) had two or more responses for one wave, and 3) took less than 600 seconds (10 minutes) to complete the survey. This cut-off point was decided based on it being close to the lowest in the distribution of response durations. This left us with a remaining sample of 364, 259 and 224. Attrition was due to difficulties with downloading the software required to play the approach-avoidance task online. Those who dropped out of the study and those that stayed in were comparable across gender, age, and income (Appendix, Tables 22-24). Finally, since we were interested in changes within people over time, we removed those that did not have data in all three waves, as well as cases with missing approach-avoidance data, which were not feasible for analysis. A final sample of 174 remained in each wave. The present research was well powered to detect at least medium effect sizes, regardless of the alpha level used (for full calculation description see Appendix, sample size calculation, page 15).

Out of the 174 participants in this study, 53% were male and 47% were female; 20% were key workers, 80% were not. There was a roughly even split across age groups (14-30% each in age brackets 25-34, 35-44, 45-54, 55-64 and 65+) although there was less representation for ages 18-24 (2%). The sample had good representation across income levels and regions. However, it lacked representation from other ethnic groups since 97% of the sample were white, which we address as a limitation of the study (see Appendix, Tables 18-21 for full breakdown).

Procedure

We administered an online survey to the same people at three different time points spaced one month apart. The online survey consisted of various survey questions and a Stimulus Response Compatibility Task (SRCT; also called "manikin task"), which measures automatic approach-avoidance tendencies. The survey questions included items on mood, wellbeing, personal traits, and adherence to 7 of the prescribed social behaviours. After answering these questions, participants completed the SRCT. During the SRCT, participants were presented with an image in the centre of the screen, and a matchstick man stood either above or below the image. Participants had to pay attention to whether the image was tilted to the right or left. They were told that hitting the B key made the man run down and hitting the Y key made the man run up. If the image was right-tilted, they were instructed to make the matchstick man AVOID the image by hitting the correct key (B or Y, depending on whether the man was located below or above the image). If the image was left-tilted, they were instructed to make the matchstick man APPROACH the image, using the same keys. When the B or Y key was pressed, a short movie was started, showing the manikin move slightly downwards or upwards, respectively. The task contained images of social scenes, as well as positive and negative facial expressions. We used participant button press reaction times to measure approach-avoidance tendencies. Full task description can be seen in Appendix, figure 1. Figure 1 below summarises the procedure.

The happy and sad face images were sourced from the Radboud Faces Database. All face images had been rated by a large pool of diverse participants according to their valence and arousal dimensions (Langner et al., 2010). We selected 8 male and 8 female actors from the Radboud Faces Database; each one expressed each emotion, yielding 16 positive (happy) facial expressions and 16 negative (sad) ones. In addition, 32 social scene images were sourced from three databases of images (Lang et al., 1999; Marchewka et al., 2014; Wessa et al., 2010) which are widely used by the academic community. Social images were diversified across image sub-types: crowds, everyday scenes, pairs, and small groups. Each participant first completed 16 practice trials, then 60 happy/sad face image trials, then 20 social image trials, yielding a total duration of approx. 10 min. For each of the 4 image types, 10 pictures had to be approached and 10 had to be avoided by the manikin. Across the 4 image types, we counterbalanced which pictures had to be approached versus avoided. From the available pictures, the program randomly picked the ones to be used for each test run.

Since it was not possible to seamlessly integrate the SCRT into the survey form, to complete the task participants were directed to a separate online link which required a software download to run the task. To match the SCRT data with the survey data, participants were required to enter a memorable word both in the survey form and on the task link before completing the SCRT. A separate file was then created containing data where we were able to achieve a match between survey and online task data and a separate file was created with these participants only for data analysis.

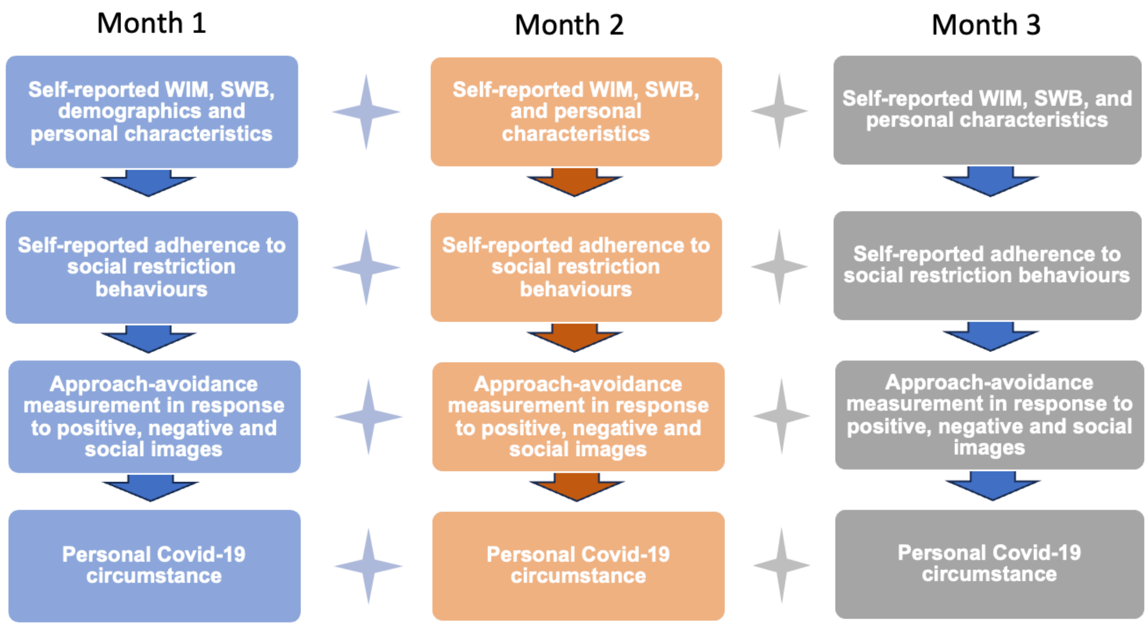


Figure 1. *A schematic diagram depicting the study procedure. The blocks detail the key data collected via online surveys at each month for the same individuals over three time points spaced one month apart. The data collected remained largely consistent at each time point, however, the static characteristics measured at time 1 (e.g. age, gender, socio-economic status etc.) were not taken at the later time points since no change was expected on these variables.*

Measures

We began the survey by asking participants to report their subjective wellbeing based on the four questions used by the Office for National Statistics in the UK: "*Overall, how satisfied are you with your life nowadays?*" measuring the evaluative dimension; "*Overall, how worthwhile are the things that you do in your life?*" measuring the eudemonic dimension; "*Overall, how happy did you feel yesterday?*" and "*Overall, how anxious did you feel yesterday?*" both measuring the affective dimension. To capture stress we added the question, "*Overall, how stressed did you feel yesterday?"*. Reports are on a 0-11 scale, from 'not at all' to 'completely'.

We then measured affect using the Affective Slider (Betella & Verschure, 2016) which allows for the relative assessment of valence and arousal on a comparative scale. This is important when trying to determine which properties of moods might be linked to different motivational tendencies.

This was followed by two questions where participants had to indicate the extent to which they agreed with the following statements: "*I am worried about catching the Coronavirus (for me)*" and "*I am worried about catching the Coronavirus (for others)*"

We then measured adherence to prescribed social behaviours (Table 1) by asking participants to indicate "the extent to which the following statements describe your behaviour over the past 7 days" on a scale of 0 (not at all) to 11 (completely).

This was followed by several self-report trait measures including the State Trait Anxiety Inventory (Spielberger, 1993) measuring generalised anxiety, the Liebowitz Social Anxiety Scale (Heimberg et al., 1999) measuring social anxiety, a brief measure of personality (Big 5, Gosling et al., 2003) and the BIS/BAS scale (Carver & White, 1994) which is designed to measure two motivational systems: the behavioural inhibition system (BIS) and the behavioural activation system (BAS), which correspond to approach-avoidance tendencies. Finally, we asked a series of questions about participants' current circumstances (e.g., COVID-19 symptoms, how many people they live with, and general health) as well as demographics. After filling out this online survey, participants completed the SCRT.

Model descriptions

Fixed effects were selected as primary models for their ability to omit unobserved variable bias. However, random effects were also conducted on significant outcomes for comparison and robustness. Since we were interested in whether changes in behaviour were significantly associated with changes in approach-avoidance tendencies within people over time, we mainly focused on participants with data points on these variables for all three waves. However, for robustness, and to deepen our understanding of the data, we also ran models on a maximum power sample, which included people with data from 2 time points in addition to those with 3 time points, and a maximum change sample which included people with data from time points 1 and 3 only, where we will have seen the biggest change in social behaviour over time. Tendency scores were calculated by subtracting the mean approach reaction times from the mean avoidance reaction times for each stimulus type: happy, sad, and social.

In all initial panel linear models, social behaviours 1-7 were entered as predictors and ID and Time were kept constant. Any identified effects of social behaviour on approach-avoidance tendency outcomes from these models were then explored further in more robust panel linear models which included all time-varying covariates.

An example fixed effects model we tested to explore the relationship between less overall interaction and sad tendency is listed below.

$$\begin{aligned} Sad\ Tendency_{it} &= \mu_t + \beta_1 behaviour6_{it} + \beta_2 Corona\ fear\ (self)_{it} \\ &+ \beta_3 Corona\ fear\ (others)_{it} + \beta_4 Anxiety_{it} + \beta_5 Social\ Anxiety_{it} \\ &+ \beta_6 Anxious_{it} + \beta_7 Happy_{it} + \beta_8 BIS_{it} + \beta_9 BAS\ Drive_{it} \\ &+ \beta_{10} BAS\ Reward_{it} + \beta_{11} Hours\ Away_{it} + \beta_{12} Stressed_{it} \\ &+ \beta_{13} Life\ Satisfaction_{it} + \beta_{14} Valence_{it} + \beta_{15} Arousal_{it} + \alpha_i + \varepsilon_{it} \end{aligned}$$

, where μ_t is the intercept term which varies across time but not cases, α_i captures the entity effects (or fixed effects) for each individual participant i, t denotes every time (wave) for each participant and ε_{it} is the error term. A two-way analysis was chosen since we were interested in both subject and time effects. All models were run using the panel linear regression model (Panel Linear Regression, PLM, 2022) package in R Studio (Version 1.4.1717) and standard errors were clustered at the level of individual i in all models.

Using the overall sample standard deviation of the Behaviour 6 (less-overall-social-interaction) coefficient, we computed Cohen's *d* to assess the magnitude of the significant effects. Data and statistical analyses are available on request.

Data availability

The datasets generated and analysed during the current study are available in The Open Science Framework repository, [https://osf.io/ydqgm/]

Personal reflections on paper 3

These are personal research reflections and do not form part of the main paper. The purpose of this section is to highlight analysis challenges and key research skills developed.

Analysis challenges

One of the biggest challenges in analysing these data was deciding whether to use a panel linear regression or a multi-level regression model. I opted for panel linear regression analyses since they are free from distributional assumptions (PLM, 2022). Multi-level models are typically used where there is substantive interest in group effects and are more common than panel linear models in psychology (University of Bristol, n.d.). However, the main point of interest here was within-person effects therefore panel linear regression was used. Moreover, the structure of these data was such that most of the variables assessed varied over three time points. In these cases, and especially since we also incorporate unbalanced models, panel linear regression models are preferable (Econometrics with R, n.d.; Croissant et al., 2008).

Another important decision that had to be made was whether the panel liner model should be a random or fixed effects model. I understand that it is best practice to select this on the basis of theoretical reasons before analysis. However, since in the context of our work this decision was not straight forward (there were no obvious theorised reasons to select one over the other) I conducted both fixed and random effects models and highlighted the model with the best fit, as indicated by the Hausman test (Amini et al., 2012).

Another important issue that arose during the data analysis for this paper was that, due to higher sample attrition than expected, the sample size was not as large as was initially hoped for. To address this shortcoming, I re-ran the main panel linear regression analyses comparing the original model with a model that includes people who have data points from two as well as three time points to achieve *maximum sample*. With this model I was able to obtain the largest possible sample, increasing the original sample from 522 to 598.

I also tested for mediation and moderation of the observed effect by certain variables. Unfortunately, there are some major flaws associated with mediation in the case of the present research design and so I was not able to include this in the analysis (Fielder et al., 2011; Fielder et al., 2018). First, all pathways are correlational, so it is impossible to argue whether behaviour causes certain emotions which in turns cause approach-avoidance tendencies, or emotions cause behaviours which in turn cause approach-avoidance tendencies, or something else. Similarly, because the study tested behaviours, approach/avoidance tendencies, and affective variables at the same time points, it is not possible to say with certainty what precedes what. To address this issue, I instead included important variables as covariates and additionally explored the association between the variables theorised as being important for the relationship between social behaviours and approach-avoidance tendencies.

Since this analysis required advanced coding skills, I sought out tuition using LSE funding to obtain the necessary advancements. These advancements included: how to specify fixed and random effects, how to convert variables to long format, how to run panel linear models as well as multi-level models (the latter were required for our revise and resubmit).

Key research skills developed

I took a lead role in this research and was responsible for all parts of the research process from study conception to publication.

This research taught me how to utilise changes in the natural environment to explore research questions that might otherwise be out of reach. The longitudinal panel study design was chosen since it suited the research question: are changes in naturally occurring WIM-related behaviours (social interaction) associated with approach-avoidance tendencies (which have been closely linked to mental wellbeing). Conducting the study at a time when social behaviours were naturally changing across the UK population also meant the ecological validity – defined as a study conducted with minimal disturbance to the contextual ecology of that setting - of our study was high (Roe & Just, 2009). The more natural ranges of treatment effects within organically formed contexts that natural experiments allow for contribute to high external validity (Roe & Just, 2009). Though, I do acknowledge that the pandemic context might also have a limiting impact on ecological validity here which I refer to in the study limitations.

I recognise that it would have been advantageous to measure WIM-related behaviours at separate time points to approach-avoidance tendencies to enable us to make claims about prediction (Kelloway & Francis, 2013). However, unfortunately time pressures and budget didn't allow for this more complex design. It was more important, given the timely nature of this study, that we conducted the study whilst lockdown measures were being eased in the UK and social behaviours were expected to change significantly.

Given the pandemic context within which this study took place a high degree of flexibility and adaptability was required. I had to quickly learn how to administer approach-avoidance measurement in an online setting and create a new set of social images to test for a potential impact of reduced social interaction on social images. I also had to overcome difficulties with linking the datasets across waves which was not possible to do automatically with the approach-avoidance software. To resolve this, I asked people to write down a memorable word and I used this word to link their datasets across waves.

3.4. Paper 4

Reporting wellbeing-in-the-moment, thoughts and context reduces anxiety

Amanda Henwood*[1], Luc Schneider*[1], Aleksandar Matic[2] and Paul Dolan[1] and Teodora Sandra Buda[2]

[1] Department of Psychological and Behavioural Science, London School of Economics and Political Science (LSE), London, United Kingdom
[2] Koa Health, Barcelona, Spain
* Joint first authors

Abstract

Increased focus on wellbeing and mental health in the past decades has made overall mental wellbeing interventions an important priority for policymakers and social scientists alike. Popular experiential subjective wellbeing (SWB) measures require individuals to repeatedly report on their wellbeing-in-the-moment (WIM). Given that this process may help to facilitate emotion differentiation and affective awareness which are associated with positive mental wellbeing outcomes, reporting on WIM and its associated context holds potential as a relatively easy and impactful mental wellbeing intervention. However, research exploring the impact of WIM-reporting on overall mental wellbeing is still in its infancy. This paper presents the results of three Randomised Controlled Trials exploring the impact of reporting WIM, thoughts, and their context, via popular SWB measures on various mental wellbeing outcomes. Across all studies, participants of diverse socio-cultural backgrounds completed daily Ecological Momentary Assessment (EMA) and/or Day Reconstruction Method (DRM) subjective wellbeing questionnaires for two to three weeks and showed medium-large sized reductions in self-reported anxiety as a result. Effect sizes were comparable to existing behavioural and positive psychology interventions, suggesting that the simple

act of reporting WIM and its surrounding context may be as effective at improving mental wellbeing as more complex and time-consuming interventions.

1. Introduction

Experiential subjective wellbeing measures require individuals to provide daily reports on their wellbeing-in-the-moment (WIM) and thus force them to pay frequent and close attention to their momentary affective states. They also often involve accompanying contextual reports such as thoughts, activities, location and company, that may help to facilitate making sense of affective experiences. Whilst the idea that paying attention to how we feel can help us to feel better overall is a long-standing one in philosophy and psychiatry (Cloniger, 2006; Pennebaker, 1997) research exploring the impact of frequent WIM reports on overall mental wellbeing is lacking.

This paper uses the term mental wellbeing to refer to both subjective wellbeing and mental health. Subjective wellbeing can be broadly defined as how people feel as they go about their daily lives. Measures of subjective wellbeing can be categorised according to whether they tap into evaluative or experiential wellbeing (Dolan et al., 1017). Evaluative measures typically ask people to aggregate their affective experience in response to a single prompt such as “overall, how satisfied are you with your life nowadays”, whereas experiential measures are an actual aggregation of multiple wellbeing reports from day-to-day moments over a period of time. Mental health is defined as the absence of mental illness, which refers to the ongoing experience of negative affective experiences linked to clinical disorders, such as worry and rumination (Lamers et al., 2011).

Given their higher potential for affective impact, considering their focus on collecting numerous WIM reports to approximate SWB, the present study focuses on the potential of experiential (as opposed to evaluative) subjective wellbeing measures to impact on overall mental wellbeing. The Ecological Momentary Assessment (EMA, Stone & Shiffman, 1994), which has been used to capture momentary affective states multiple times a day, and the Daily Reconstruction Method (DRM, Kahneman et al., 2004), a one-off record of retrospective accounts of affective moments from the previous day, are two of the most common ways of capturing subjective wellbeing. When being used to approximate subjective wellbeing, EMA and DRM typically involve asking people to report on their wellbeing in the moment (WIM), thoughts, and the surrounding context.

Interpersonal theory has highlighted interpersonal situations as important determinants of mental health (Pincus & Ansell, 2013). Interpersonal situations refer to contexts where it is necessary for people to relate to other people or things in their environment. According to interpersonal theory, if a person's needs and their context (e.g. the activity they are engaged in) are not complementary then they are likely to experience a negative affective response. Requiring people to record the contextual factors associated with their affective experiences may help people to better attend to and therefore understand their patterns of interpersonal relations, resulting in higher emotional awareness. There is also extensive evidence that both thoughts and activities are strongly related to wellbeing (Killingsworth & Gilbert, 2010; Smallwood & Schooler, 2015; Stawarczyk et al., 2012; White & Dolan, 2009). It follows, therefore, that contextualising wellbeing reports by accompanying them with thoughts and external context reports may be particularly impactful.

Factoring in this research, we recognise two potential reasons why reporting on how we feel, what we think about and what we do, might impact how we feel more generally. First, reporting on WIM, thoughts and their contexts may help to increase affective awareness. Higher emotional self-awareness has been shown to be an important step in addressing mental illness symptomology (Kauer et al., 2012). It has been suggested that awareness of one's thoughts and feelings facilitate emotional stability by shifting attentional focus away from the direct experience of these feelings and promoting emotional acceptance (Coffey et al., 2010; Fogarty et al., 2015; Frewen et al., 2008). Interventions inducing such kinds of awareness have been shown to encourage reflection through identification of thoughts and feelings without judgement (Louet, 2015). They have also been shown to produce improvements on a variety of mental wellbeing indices including anxiety and depression (Enkema et al., 2020; Hofmann et al., 2010). Relatedly, there is a compelling body of evidence showing that poor interoceptive awareness is associated with emotion regulation difficulties (Price & Hooven, 2018; Edwards & Wupperman, 2017; Demiralp et al., 2012).

Second, identifying and distinguishing between different feelings is thought to help people better respond to those experiences, since it enables a more specialised and therefore adaptive response (Kashdan et al., 2015). For example, people who experience their emotions with higher granularity, i.e., using more words to describe both positive and negative emotions, are less likely to use maladaptive coping strategies such as aggression and self-medicating (Kashdan et al., 2010) and are more likely to use positive emotion regulation strategies that target the specific emotion (Barrett et al., 2001). Self-monitoring of emotions using a mobile app over a 6-week period has been shown to have positive impact emotion differentiation in depressed individuals (Widdershoven, 2019). Interventions that train people to improve their emotion differentiation before aversive experiences have also been shown to reduce anxiety during those experiences more than other popular interventions such as cognitive reappraisal and distraction (Kircanski et al., 2012).

Most of the research on EMA reactivity effects to date has focused on behavioural as opposed to affective reactions. For example, studies exploring the impact of EMA reports on alcohol intake have revealed mixed results, with some research identifying positive associations between EMA reports and alcohol intake reduction (Collins, 1998), and others finding no such associations (Hufford et al., 2002; Litt et al., 1998). Research exploring the impact of EMA reports on body-image related concerns and eating-disordered behaviours have found no association (Heron & Smyth, 2013; Stein & Corte, 2003). With respect to smoking, one study found no impact of high EMA report frequency or low EMA report frequency on smoking-related behaviours (McCarthy et al., 2015). However, it did find a strong association between high EMA frequency with lower craving and anger. Similarly, another study has identified no change in perceived risk of smoking but did observe less reported worry, suggesting that the impact of EMA reporting on affective components warrants further exploration (Magnan et al., 2013).

Studies in the clinical literature that have explored the impact wellbeing reporting on mental health have shown some promise. For example, studies assessing momentary wellbeing alongside contextual reports such as current activities and companions (or feedback encouraging people to consider the contexts surrounding their emotional reports) have found that EMAs increase people's awareness of how they feel (Widdershoven et al., 2019; Kauer et al., 2012) and that this awareness may help to prevent depression or function as an early intervention against it (Beames et al., 2021; Kauer et al., 2012; Kramer et al., 2014). One such study found that providing face-to-face feedback on how patients' affect and depression levels were related to their daily activities produced even stronger, longer-term reductions in depression, suggesting that context matters. However, another study that got participants to engage in symptom tracking (e.g. anxiety, depression, irritability, restlessness, stress and worry) alongside context reports (location, activities and company) as part of a broader intervention programme found no significant association between symptom tracking and context reports with post-intervention anxiety, stress and functional impairment. Importantly, this body of research is either limited to a focus on one aspect of mental wellbeing only (e.g. depression), on negative WIM reports only, or on assessing the impact of EMA in combination with other interventions which poses significant challenges for effect isolation (Myin-Germeys et al., 2018; Myin-Germeys et al., 2016; Balaskas et al., 2021; Schueller et al., 2017).

There is a dearth of research exploring the impact of EMA and DRM reactivity effects in subjective wellbeing literature. However, one study has found that reporting happiness levels several times a day for two weeks accentuates pre-existing levels of momentary happiness– namely, people with relatively poor mental health were worse off, while people with better mental health were better off (Conner & Reid, 2012). Notably, this research did not include thoughts and context (i.e., activities, company, and location) alongside WIM reports, and so it might *not* have allowed participants to reflect more broadly on their day-to-day life and contextualise their feelings.

Overall, the impact of contextualised WIM reports on different facets of mental wellbeing (e.g. subjective wellbeing in addition to mental health) remains insufficiently explored, particularly across different mental wellbeing outcomes, time points and cohorts.

2. The present work

To address this gap in research, we conducted three randomised controlled trials (RCT) in which we compared individuals who completed EMA/DRM questionnaires for two to three weeks with individuals who completed the same questionnaires removing the wellbeing related items (studies 1 and 2), and also individuals who did not complete any EMA/DRM questionnaires during the two to three week period (study 3). We assessed the impact of WIM reporting in the different groups on mental wellbeing outcomes. This research is the first of its kind to evaluate whether the simple act of reporting WIM and the context within which it is experienced affects overall mental wellbeing outcomes spanning clinical *and* non-clinical outcomes: including life satisfaction, happiness, worthwhileness, daily anxiety, daily sadness, clinical anxiety, clinical depression, and stress. This makes an important contribution to intervention science literature, which has called for more research exploring the impact of specific intervention components on psychological outcomes (Patel et al., 2018; Holmes et al., 2014; Firth et al., 2017; Sin & Lybumirsky, 2009).

Hypotheses for studies 2 and 3 were pre-registered as part of a broader set of predictions. The pre-registered hypothesis for study 2 was "asking about subjective wellbeing every day for 2-3 weeks will change overall subjective wellbeing" (Open Science Framework, 2019). The pre-registered hypothesis for study 3 was "after 2 weeks people in the treatment group will have a higher average change in outcome measures than the control group 2" (Open Science Framework, 2021). Study 1 was conducted before pre-registration practices were established in our working practice, however, the hypothesis we had for this study was the same as that in study 2.

All studies reported in this paper were approved by the LSE Ethics Committee and conducted in accordance with APA ethical guidelines. Active, informed consent was collected from all participants prior to each study, and participants were able to request retraction of their personal data at any time following the study. We report all manipulations, measures, and exclusions, and no data collection took place after any stage of data analysis.

Analysis plan

Prior to analysis, we determined identical exclusion criteria for participants. Since participants across all studies were asked to fill in DRM questionnaires (i.e., 21 in study 1, 14 in studies 2 and 3, one for every day over a two-week time period), we used the number of DRMs filled in as a benchmark for whether or not individuals could be said to have received the "treatment". As such, we excluded all participants who filled in less than five DRM questionnaires, as well as those who did not complete the onboarding and exit questionnaires. To avoid excluding people from study 1 and 2 who satisfied the inclusion criteria of study 3, we did not set a minimum number of completed EMAs.

Between group differences: comparing questionnaire delta scores (onboarding minus exit questionnaire scores) of the treatment and control groups

We first assessed the impact of the treatment on mental wellbeing outcomes by comparing the deltas between onboarding and exit questionnaire values of each mental wellbeing measure for the treatment and control groups, in each study. Since distributions of subjective wellbeing reports and mental illness measures tend to be skewed (towards more positive outcomes), we used non-parametric tests (Mann-Whitney U).

Within group differences: exploring significant effects by comparing change scores on the relevant outcome variables for treatment and control groups, respectively

We then looked at within group differences in onboarding and exit questionnaires for all outcome variables for any significant effects identified. This follow-up analysis allows us to capture the direction and magnitude of any changes in mental wellbeing outcomes between the start and end of the study in both control and treatment groups. This is important to determine where any identified effects come from (e.g., whether an effect is driven by increases in wellbeing in the treatment, or reductions in wellbeing in the control group).

We report two-sided p-values throughout the manuscript, to account for the fact that we did not specify any hypotheses in study 1. As these yield more conservative estimates of significance and studies 2 and 3 focus on confirming the findings of study 1, we elected not to use multiple hypotheses adjustments. Using the overall sample standard deviation of each wellbeing and mental health measure, we computed Cohen's *d* to assess the magnitude of the significant effects.

Finally, we ran simple linear regressions using the number of EMA and DRM questionnaires answered as explanatory variables to predict changes in the relevant mental wellbeing measures in the treatment group. These models are used to check whether answering more questionnaires is indeed associated with greater improvements in wellbeing, as should be the case for all mental wellbeing variables for which the treatment was effective.

3. Study 1

In study 1, participants were asked to download an app, which randomly assigned them to either a treatment group, or an active control group receiving a sham EMA and DRM in which momentary wellbeing reports were excluded. All participants were asked to complete up to five real/sham Ecological Momentary Assessments (EMA) and one Day Reconstruction Method questionnaire (DRM) every day for three weeks.

Method

Procedure. We asked adults in four Spanish-speaking countries (Spain, Chile, Columbia and Peru) and the UK to download an app called Reflections, a research curated app designed for the purposes of this study. The study took place between February and August 2018. Participants first filled in an onboarding questionnaire containing demographics and trait-based questionnaires and the key outcome measures (life satisfaction, worthwhileness, happiness yesterday, anxiety yesterday, happiness in general and anxiety in general). One third of participants were randomly sorted within the app into the control and two thirds into the treatment group. Participants were notified between five evenly spaced windows throughout the day to answer one EMA (five per day), and one DRM every day, over the course of three weeks. In the treatment group, participants answered questions related to their WIM and its context (thoughts, activities, location and company). In the control group, participants answered the same questions, except for those relating to WIM, which were replaced by random questions asking them to rate something on a 0-10 scale (e.g., “How much do you like reading books?”). Upon completion of the study, participants were prompted to answer a final survey asking them to report on the same mental wellbeing measures as in the onboarding survey.

Participants were paid €40 (or $45) upon completion of the study. They were considered to have completed the study if they answered at least 80% of questionnaires and filled in both the onboarding and exit questionnaires. Figure 1 below summarises the procedure for each of the 3 studies. Further detail on each study can be found in the respective study sections.

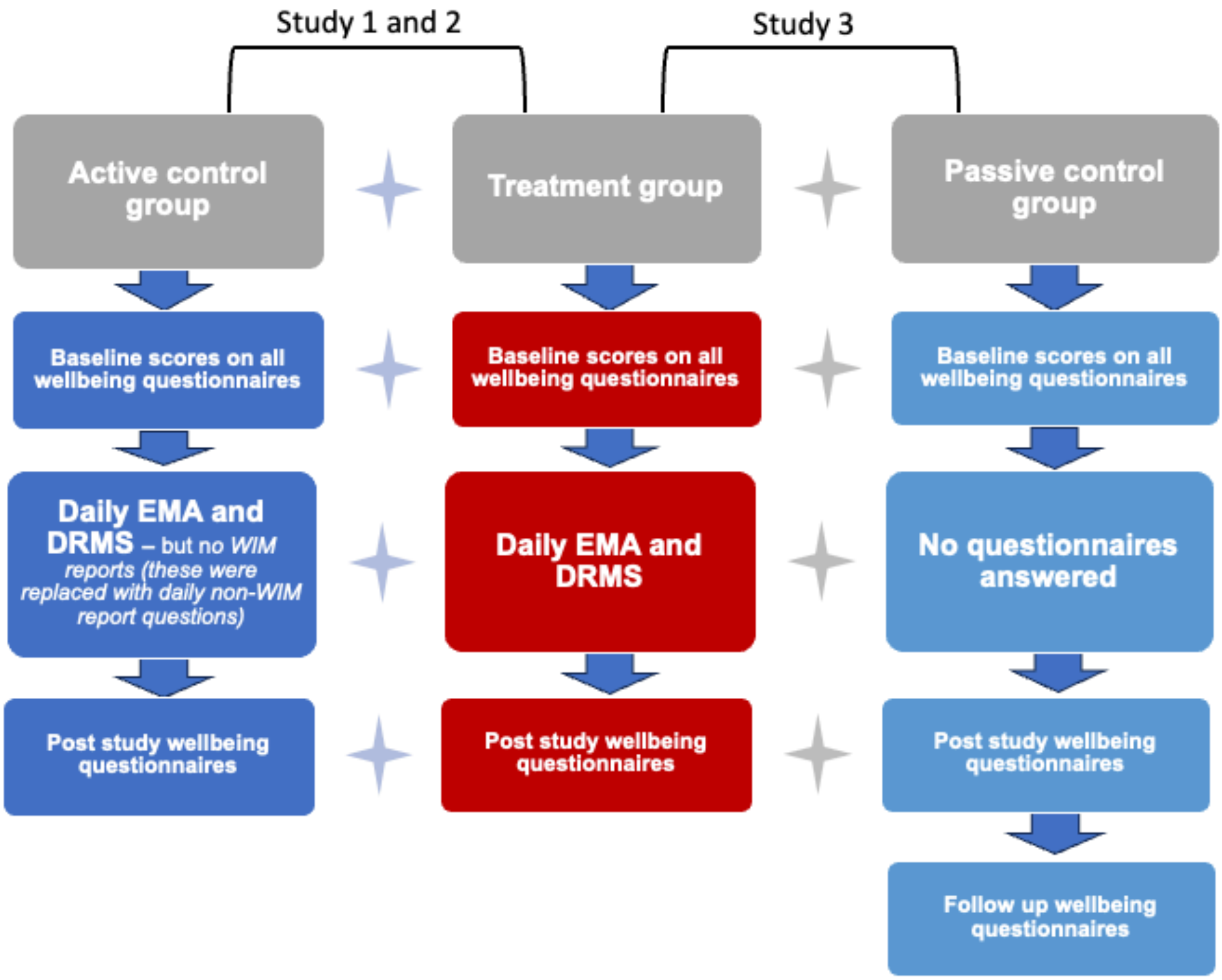


Figure 1. *The flow of procedure for all three studies. The main differences occur in the type of daily questionnaires that participants filled out: daily EMAs and DRMs without an affective component, daily EMAs and DRM with an affective component, and no EMAs or DRMs at all.*

Participants. Across all studies, we only included consenting adults (over the age of 18). In study 1, 691 participants completed the onboarding survey and were allocated to the control and treatment groups. Of these, 396 participants completed at least five DRM questionnaires (either sham or real), as well as the onboarding and exit questionnaires. As such, our final sample is composed of 123 participants in the control group, and 273 participants in the treatment group. Of these, the majority were from Spain (47.0%), and 8.3% were from the UK. The rest were from Chile, Columbia and Peru. 36.1% of the sample were female, 30.6% were students, and 43.9% were employed. 68.2% of participants reported being between 26 and 34, while 11.1% were younger, and 18.2% reported being between 35 and 44.

Daily reports and experiential subjective wellbeing measures

Ecological Momentary Assessment. EMA (Stone & Shiffman, 1994) reports consisted of responses to various prompts issued at five random intervals throughout the day. First, participants had to select an activity (e.g., “working”) from a list of common activities in response to the prompt: “During the past hour, I was”. Activity lists differed depending on whether the sample was student or mixed. Student samples received common activities (e.g., “eating”) in addition to activities that were tailored to university life (e.g. “studying”). Non-student samples only received common activities. Next, participants had to indicate the duration of the activity so far using a drop-down tab which showed time periods that went up in 10-minute increments, ranging from 10 minutes to 4 hours and 10 minutes, in response to the prompt: “How long have you been doing this?”. Then, participants had to indicate who they were with, what they were thinking about, and where they were, from a list of common suggestions (e.g. “Kids”, “Events from my past”, “At my parents’ house”) in response to prompts: “I was with”, “I was thinking about”, “Where are you?”. Finally, participants had to report their momentary wellbeing on a scale of 0-10 in response to prompts: “How happy did you feel?” and “How worthwhile did this feel?”. For the active control group, the final momentary wellbeing and thought reports were removed. The timing of EMAs was randomised.

Day Reconstruction Method. Every morning, participants were asked to provide an overview of the previous day partitioned into episodes. We used the text from the DRM instructions provided in Kahneman et al, 2004: "Think of your day as a continuous series of scenes or episodes in a film. Give each episode a brief name that will help you remember it (for example, `commuting to work', or `at lunch with B'...). Write down the approximate times at which each episode began and ended."

Participants then had to indicate what they were doing (e.g. "working") in response to the prompt "I was doing" and select a "start time" and "end time" for the episode. Activity lists differed depending on whether the sample was student or mixed, as described above in EMA reports. Also, like EMA -reports, for each episode participants had to indicate who they were with and what they were thinking about in response to prompts: "I was with" and "I was thinking about". Finally, as per EMA reports, participants had to answer how they felt on a scale of 0-10 in response to prompts: "How happy did you feel?" and "How worthwhile did this feel?" Participants had to report at least 12 hours of emotional episodes to complete the DRM.

For the active control group, the final momentary wellbeing and thought reports were removed.

Outcome measures

For the purpose of this paper, we focus only on the wellbeing and mental health measures that participants answer in the onboarding and exit questionnaires of each study. Measures common to all studies are the wellbeing questions as formulated in UK national surveys administered by the ONS (Office for National Statistics, 2019).

Evaluative Subjective wellbeing (including WIM) using ONS4. These consist of four questions asking about general life satisfaction, general sense of worthwhileness, happiness yesterday and anxiety yesterday. These questions are complemented by two additional ones asking about happiness and anxiety in general. The questions relating to happiness and anxiety yesterday are taken to be more WIM questions than subjective wellbeing questions given the recency of these experiences. All questions are answered on a scale of 0 to 10, and can be found in table 1 below.

Table 1: ONS-4 subjective wellbeing questions, extended to include questions about general happiness and anxiety

ONS wellbeing questions
Overall, how satisfied are you with your life?
Overall, to what extent do you feel the things you do in your life are worthwhile?
Overall, how happy did you feel yesterday?
Overall, how anxious did you feel yesterday?
Overall, how happy do you feel in general?
Overall, how anxious do you feel in general?

Results

Descriptive statistics

In table 2 below, we report the mean and standard deviation for each measure of mental wellbeing that participants answered in the onboarding survey.

Table 2: Means and standard deviations for each measure of mental wellbeing that participants answered in the onboarding survey. Two-sided p values reported: NS: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

	Treatment mean (SD)	Control mean (SD)	Significance of diff.
Life satisfaction	7.03 (1.63)	7.15 (1.67)	NS
Worthwhileness	7.26 (1.73)	7.37 (1.71)	NS
Happiness yesterday	6.95 (2.04)	7.43 (1.93)	*
Anxiety yesterday	5.56 (2.62)	5.21 (2.82)	NS
Happiness in general	7.22 (1.62)	7.29 (1.66)	NS
Anxiety in general	5.71 (2.32)	5.22 (2.37)	NS

We checked for significant differences between onboarding starting values of each statistic between the control and treatment group. We found that these groups did not significantly differ in terms of starting point for five of the six ONS measures. Reports of happiness yesterday appeared to be significantly higher in the control compared to the treatment group ($p = 0.030$).

Main results

Between group differences: comparing questionnaire delta scores (onboarding-exit questionnaire scores) of the treatment and control groups

As shown in Figure 2, a significant impact of the treatment condition as compared with the control was identified for anxiety yesterday and anxiety in general (anxiety *yesterday*, difference in improvement between treatment and control group = -0.96, p = 0.002; and anxiety *in general*, diff. = -0.70, p = 0.014). The size of the improvement in anxiety *yesterday* (Cohen's *d* = 0.36) and anxiety *in general* (*d* = 0.30) due to the treatment is large (see also, Funder & Ozer, 2019) However, there were no significant reductions in life satisfaction (diff. = 0.14, p = 0.253), worthwhileness (diff. = -0.06, p = 0.928), happiness *yesterday* (diff. = 0.53, p = 0.071) or happiness *in general* (diff. = 0.07, p = 0.371). See Figure 1.

Within group differences: exploring significant effects by comparing change scores on the relevant outcome variables for treatment and control groups, respectively

Specifically, we found that anxiety *yesterday* was significantly reduced in the treatment group (diff. = -0.70, p = 0.004), but not in the control group (diff. = 0.26, p = 0.423). Similarly, participants in the treatment group reported a reduction in anxiety *in general*, albeit an insignificant one (diff. = -0.44, p = 0.061), while participants in the control group reported an insignificant increase in anxiety *in general* (diff. = 0.26, p = 0.453).

In addition, simple linear regressions showed that in the treatment group, filling in more EMA or DRM questionnaires was significantly associated with greater reductions in anxiety yesterday (EMA: b = -0.008, SE = 0.002, $p < 0.001$, $r^2 = 0.056$; DRM: b = -0.008, SE = 0.002, $p < 0.001$, $r^2 = 0.049$) and anxiety in general (EMA: b = -0.005, SE = 0.002, p = 0.005, $r^2 = 0.028$; DRM: b = -0.006, SE = 0.002, p = 0.004, $r^2 = 0.031$) between onboarding and exit surveys.

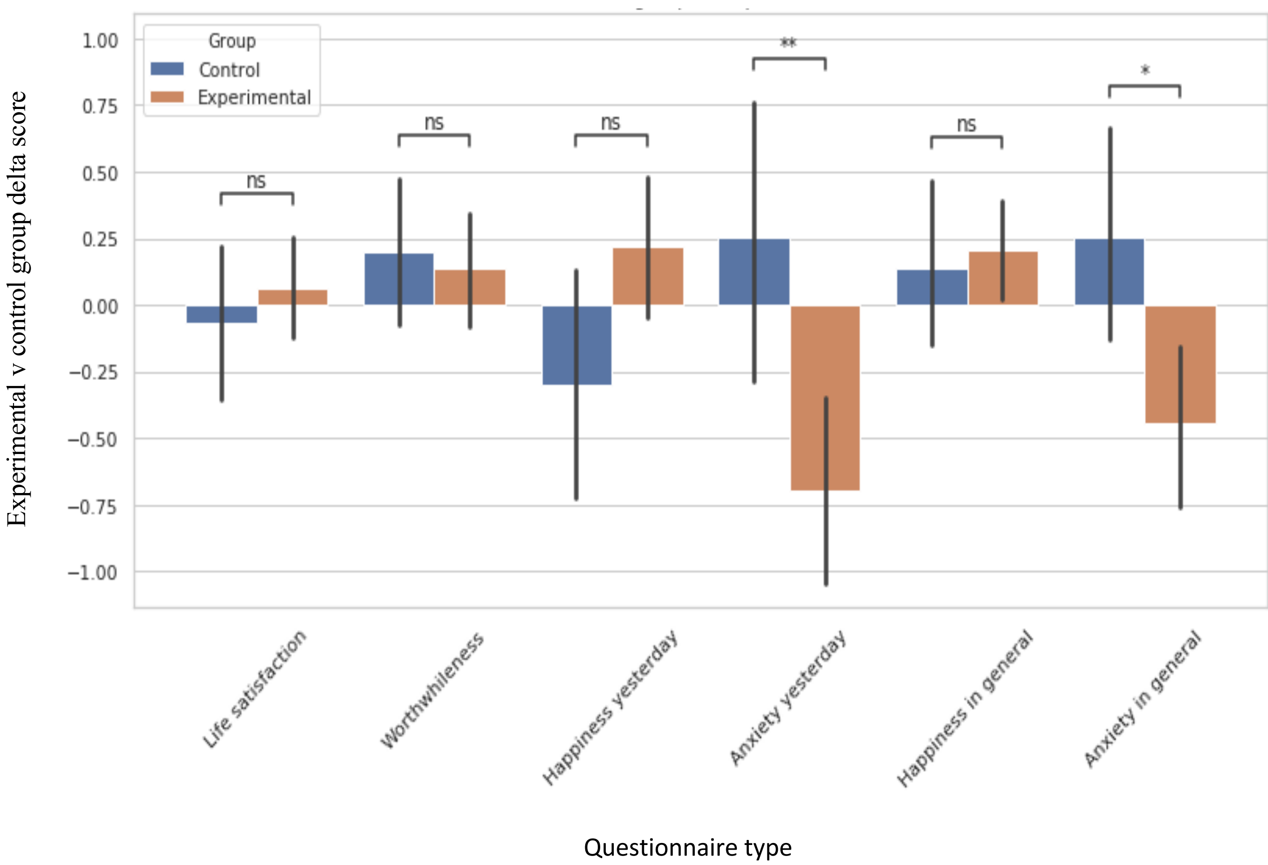


Figure 2. *Graph comparing questionnaire delta scores between experimental and control conditions for each outcome variable. Statistical significance was calculated using Mann Whitney U tests to compare delta (onboarding-exit) scores between experimental and control groups for each outcome variable. Two-sided p values reported: NS: not significant, *: p < 0.05, **: p < 0.01, ***: p < 0.001.*

Discussion

Study 1 revealed large and significant between-group effect sizes for reductions in anxiety between the treatment and control groups on both anxiety yesterday and anxiety in general outcome measures. This suggests that reporting on WIM and its associated context might have contributed to reductions in short-term and general anxiety.

Study 1 also revealed significant within-person reductions for anxiety yesterday (using Mann-Whitney U tests and regression analyses) and for anxiety in general (using the regression analyses only) within the treatment group, but not in the active control group. However, we do note that the reduction in anxiety in general was approaching significance ($p=0.061$). Therefore, it appears that, overall, reporting how you feel alongside your current context has potential as an effective intervention for reducing anxiety levels.

Study 1 did not identify any significant improvements in life satisfaction, worthwhileness, happiness yesterday or happiness in general, suggesting that this intervention is more effective for anxiety than the other measured affective states.

4. Study 2

Similar to study 1, in study 2 participants were asked to download the same app which randomly assigned them to either a treatment group, or an active control group receiving a sham EMA and DRM in which momentary wellbeing reports were replaced with random questions. Participants were asked to complete up to five real/sham Ecological Momentary Assessments (EMA) and one Day Reconstruction Method questionnaire (DRM) every day this time for two weeks. Instead of the general population across multiple counties, this study was targeted at UK university students from the London School of Economics.

Method

Procedure. Data for study 2 were collected between January and February 2019. The procedure for this study was the same as study 1, except that this time, only around one quarter of participants were randomly allocated to the active control group in which all WIM-related questions were replaced with random questions (e.g., "How much do you like reading books?").

Participants were paid £20 upon completion of the study. They were considered to have completed the study if they answered at least 70% of questionnaires and filled in both the onboarding and exit questionnaires. Completion requirements were adjusted downward after the first study to avoid excessive drop-outs from people who had already missed a few questionnaires early on in the study. In addition, participants who filled in at least the onboarding survey were paid £5 for partial completion, if they did not meet the required completion threshold.

Participants. In total, 666 LSE students filled in the onboarding survey, of which 348 satisfied the inclusion criteria as previously specified. The final sample was therefore composed of 93 participants in the control group and 255 in the treatment group. 64.7% of the final sample were female, and 15.5% reported being employed. Most participants were younger than 26 (82.6%).

Measures. Measures used were the same as those in study 1.

Results

Descriptive statistics

In table 3 below, we show the mean and standard deviation for all mental wellbeing measures in the treatment and control groups, respectively, in the onboarding questionnaire. In study 2, there were no significant differences in the wellbeing levels reported by people in either group. Similarly, there were no significant differences in the gender, age and employment distributions across these groups.

Table 3: Means and standard deviations for each measure of mental wellbeing that participants answered in the onboarding survey.
*Two-sided p values reported: NS: not significant, *: p < 0.05, **: p < 0.01, ***: p < 0.001.*

	Treatment mean (SD)	Control mean (SD)	Significance of diff.
Life satisfaction	6.91 (1.54)	6.79 (1.57)	NS
Worthwhileness	6.86 (2.10)	7.08 (1.77)	NS
Happiness yesterday	6.59 (1.98)	6.59 (2.03)	NS
Anxiety yesterday	5.20 (2.57)	4.86 (2.63)	NS
Happiness in general	6.78 (1.63)	6.78 (1.60)	NS
Anxiety in general	5.52 (2.33)	5.17 (2.33)	NS

Main results

Between group differences: comparing questionnaire delta scores (onboarding-exit questionnaire scores) of the treatment and control groups

A significant impact of the treatment condition as compared with the control was identified for anxiety in general (diff. = -0.42, p = 0.037). The Cohen's *d* (0.18) indicates that the size of the improvement in anxiety *in general* resulting from the treatment was medium (Funder, 2019). Whilst a reduction was observed for anxiety *yesterday,* this difference was not significant (diff. = -0.26, p = 0.606). Consistent with study 1, we found no significant effects of the treatment on life satisfaction (diff. = 0.32, p = 0.109), worthwhileness (diff. = -0.11, p = 0.821), happiness *yesterday* (diff. = 0.54, p = 0.081) or happiness *in general* (diff. = 0.24, p = 0.177). See Figure 2.

Within group differences: exploring significant effects by comparing change scores on the relevant outcome variables for treatment and control groups, respectively

Specifically, we found that the significant reduction in anxiety in general was driven by a significant decrease in reported anxiety *in general* between the onboarding and the exit questionnaire in the treatment group (diff. = -0.50, p = 0.017), while the control group reported general anxiety levels that were not significantly different between the start and the end of the study (diff. = -0.08, p = 0.971).

Our simple linear regression revealed that answering a higher number of EMA or DRM questionnaires was associated with greater reductions for anxiety in general in the treatment group (EMA: b = -0.008, SE = 0.002, p = 0.001, r^2 = 0.040; DRM: b = -0.012, SE = 0.003, p < 0.001, r^2 = 0.063) however, these differences were non-significant for anxiety yesterday.

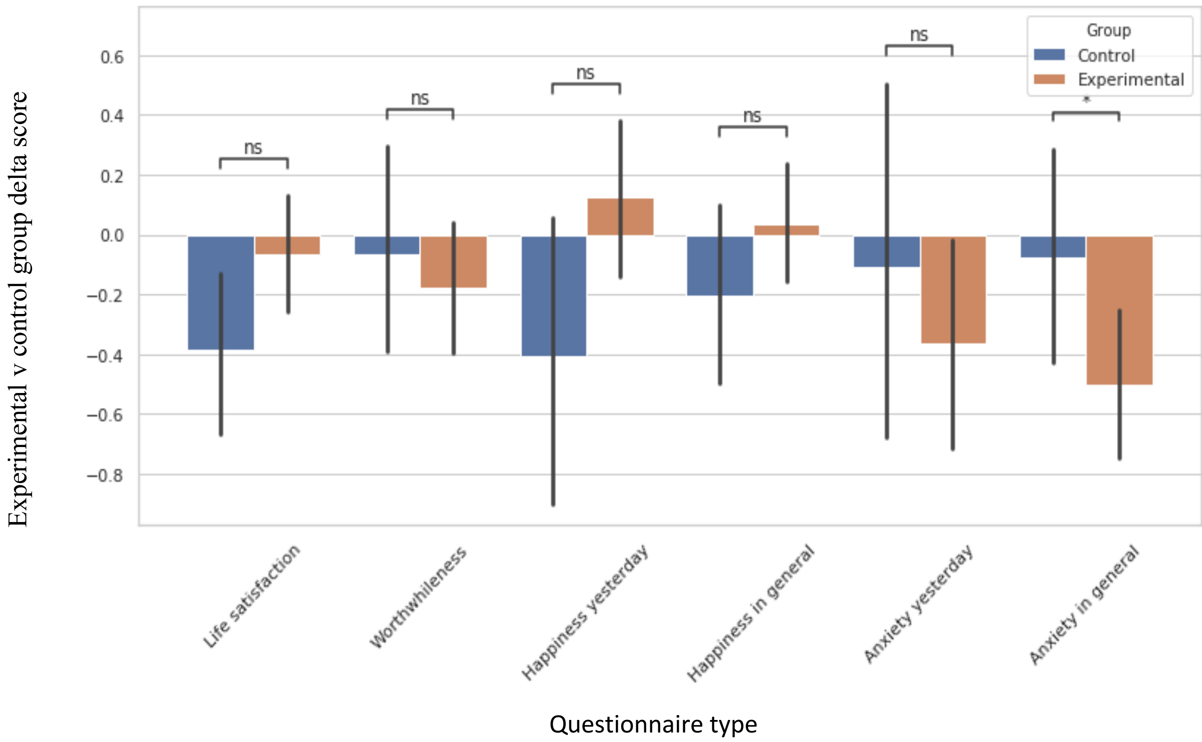


Figure 2. *Graph comparing questionnaire delta scores between experimental and control conditions for each outcome variable. Statistical significance was calculated using Mann Whitney U tests to compare delta (onboarding-exit) scores between experimental and control groups for each outcome variable. Two-sided p values reported: NS: not significant, *: p < 0.05, **: p < 0.01, ***: p < 0.001.*

Discussion

Study 2 partly replicated the positive impact of reporting on WIM alongside contextual reports on anxiety. The between-group comparisons showed a significant, medium sized reduction in anxiety in the treatment group relative to the control group for anxiety in general. However, unlike study 1, no significant reduction was identified for anxiety yesterday.

Again similar to study 1, within-person analyses revealed a significant effect of the intervention on anxiety in general in the treatment group but not the control group, suggesting once again that reporting WIM and its associated context repeatedly over the course of two weeks could help people reduce their general anxiety levels. However, unlike study 1 no significant reductions in anxiety yesterday were observed in the treatment group.

Regression analyses also highlighted a significant association between increases in EMA and DRM reports, with reductions in general anxiety levels. The association between EMA and DRM reports with anxiety yesterday remained insignificant. Therefore, whilst we observed a consistent pattern of anxiety reduction in response to repeatedly reporting WIM and its associated context across studies 1 and 2, the relative impact on momentary anxiety appears to be changeable. This may be attributed to the fact that anxiety yesterday is a more momentary measure than general anxiety, and thus more subject to fluctuation. Nonetheless, study 2 has shown that the impact of WIM-reporting on general anxiety is replicable using a different sample, and when applying the treatment for 2, rather than 3 weeks.

Consistent with study 1, we did not observe significant improvements in life satisfaction, worthwhileness, happiness yesterday or happiness in general, further suggesting that this intervention is especially effective for anxiety over and above other affective states and traits.

5. Study 3

In study 3, participants were sorted into four possible groups, including a DRM group, a passive control group and two treatments unrelated to the purpose of the present study. In the DRM group, participants were asked to download an adaptation of the app used in studies 1 and 2 that did not include EMA reports. The passive control group was asked to fill in only the onboarding questionnaire and exit questionnaires. The exit questionnaires were taken after 4 weeks (two weeks after the intervention had completed). The additional two treatments unrelated to the present study involved the use of an additional app, which served as a separate wellbeing intervention. Participants in these treatments were excluded from this study.

In addition to the ONS-4 questions that were also collected in studies 1 and 2, several scale-based measures of mental wellbeing were added to the onboarding and exit questionnaires in study 3. These additional mental wellbeing measures included trait stress, trait anxiety, trait depression, evaluative subjective wellbeing and sleep. Each scale is briefly detailed in the methods below. Like study 2, this study was targeted at university students from the London School of Economics.

Method

Procedure. This study was conducted between March and April 2021. We asked LSE students to download a new app called LSEasy, which notified them once a day to complete a DRM. Rather than downloading the app, the control group only completed the onboarding and the exit questionnaires, which were sent outside of the app. There were two other treatments in the study, in which people were asked to download a second app, called Foundations, that provided participants with mental wellbeing interventions. All participants completed onboarding and exit questionnaires consisting of the key mental wellbeing indices.

Participants were paid £30 for successfully completing the study. Those that had answered at least 70% of questionnaires and filled in both the onboarding and exit questionnaires were considered to have successfully completed the study.

Participants. In total, 610 LSE students filled in the onboarding survey, of which 306 were sorted either in the DRM treatment or the passive control group. Of these, 214 satisfied the inclusion criteria as previously specified. The final sample was therefore composed of 129 participants in the control group and 85 in the treatment group. 68.7% of the final sample were female, and 24.8% reported being employed. Most participants were younger than 26 (83.7%).

Measures. Measures were the same as those used in study 1 and 2, excluding the EMA measure, and with the additions below.

GAD-7. The Generalised Anxiety Disorder Assessment is a 7-item scale that assesses anxiety levels in participants in a more detailed manner than the ONS-4 anxiety questions (Spitzer et al., 2006). Participants are asked to rate items based on the following question: "Over the last week, how often have you been bothered by the following problems?" Items all focus on negative, anxiety-related emotions (e.g., "Feeling nervous, anxious, or on edge") and are rated on a 4-point scale of 0 ("Not at all sure") to 3 ("Nearly every day"). Scores are added up to reflect overall anxiety levels. Typically, scores of 5-9 are associated with mild anxiety, 10-14 with moderate anxiety and 15-21 with severe anxiety.

PSS-10. The Perceived Stress Scale is a 10-item scale that assesses stress levels in participants (Cohen et al., 2010). Items ask participants how often they feel certain stress-related emotions about various aspects of their day-to-day lives (e.g., "In the last week, how often have you been upset because of something that happened unexpectedly?"). The scale is composed of six negative items and four positive ones, which participants were asked to rate on a 5-point scale of 0 ("Never") to 4 ("Very often"). Positive items are reverse-coded, and scores are added up to reflect overall stress levels. Conventionally, scores of 0-13 indicate low stress, 14-26 moderate stress, and 27-40 high perceived stress levels.

WHO-5. The World Health Organisation-Five Well-Being Index is a 5-item scale that assesses general wellbeing by focusing on positive emotions (Topp et al., 2015). The questionnaire asks participants to indicate how often they have been feeling a certain way over the past week (e.g., "I have felt cheerful and in good spirits"). Participants are asked to indicate this using a 6-point scale of 0 ("At no time") to 5 ("All of the time"). The score is computed by adding up all answers. Scores of 12 or less out of 25 are understood to be indicators of poor wellbeing.

PHQ-8. The Patient Health Questionnaire depression scale is an 8-item scale that was designed to diagnose depression and assess its severity (Kroenke & Spitzer, 2002). It asks people: "Over the last two weeks, how often have you been bothered by any of the following problems?", and items focus on negative feelings associated to day-to-day activities (e.g., "Little interest or pleasure in doing things?"). Participants rate each item on a 4-point scale analogous to the GAD-7 one (from 0, "Not at all sure" to 3, "Nearly every day"). The final score is computed by adding up the rating of all items. Scores of 5–9 are considered indicators of mild depression, 10–14 moderate depression, 15–19 moderately severe depression and 20-24 severe depression.

Results

Descriptive statistics

Table 4 below shows the means and standard deviations for the mental wellbeing measures that participants reported in the onboarding questionnaire. We find no significant differences in mean reported between treatment and control groups for any of these measures. Similarly, there were no significant differences in age, gender and employment distributions between the two groups.

Table 4. Mean and standard deviation reports of mental wellbeing reports on ONS, GAD-7, PSS-10, PHQ-8 and WHO-5 measures in treatment and control groups in the onboarding questionnaire of study 3, including significance of difference based on Mann-Whitney U tests. Two-sided p values reported: NS, not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

	Treatment mean (SD)	Control mean (SD)	Significance of diff.
Life satisfaction	6.49 (1.78)	6.50(1.79)	NS
Worthwhileness	6.64 (1.87	6.63 (1.87)	NS
Happiness yesterday	6.08 (1.96)	5.94 (2.38)	NS
Anxiety yesterday	5.72 (2.20)	5.56 (2.41)	NS
Happiness in general	6.52 (1.56)	6.19 (1.61)	NS
Anxiety in general	5.74 (2.08)	5.61 (2.16)	NS
GAD-7	8.70 (5.56)	8.35 (5.26)	NS
PSS-10	22.90 (3.77)	22.23 (3.89)	NS
PHQ-8	9.58 (5.31)	9.20 (5.75)	NS
WHO-5	11.04 (5.15)	10.77 (4.64)	NS

Main results.

Between group differences: comparing questionnaire delta scores (onboarding-exit questionnaire scores) of the treatment and control groups

We found no significant effects of the treatment on any of the ONS4 questions reported in studies 1 and 2 after four weeks. We also found no significant differences between treatment and control group for measures of stress (PSS-10, $p = 0.446$), depression (PHQ-8, $p = 0.169$), or evaluative wellbeing as measured by the WHO-5 ($p = 0.715$). However, we did find that the treatment caused a significant decrease in reports of clinical anxiety relative to the control (measured with GAD-7, diff. = -1.59, $p = 0.012$). This overall impact of the treatment on clinical anxiety is of a similarly large magnitude to the effect reported for anxiety *in general* in study 1 ($d = 0.30$). See Figures 3 and 4.

Within group differences: exploring significant effects by comparing change scores on the relevant outcome variables for treatment and control groups, respectively

As in studies 1 and 2, the significant effect of the treatment on clinical anxiety in study 3 (as measured with GAD-7) was driven by a significant decrease in reported clinical anxiety scores between onboarding and exit questionnaires in the treatment group (diff. = -1.87, p = 0.027). In contrast, there was no significant decrease in anxiety reports in the control group (diff. = -0.28, p = 0.621).

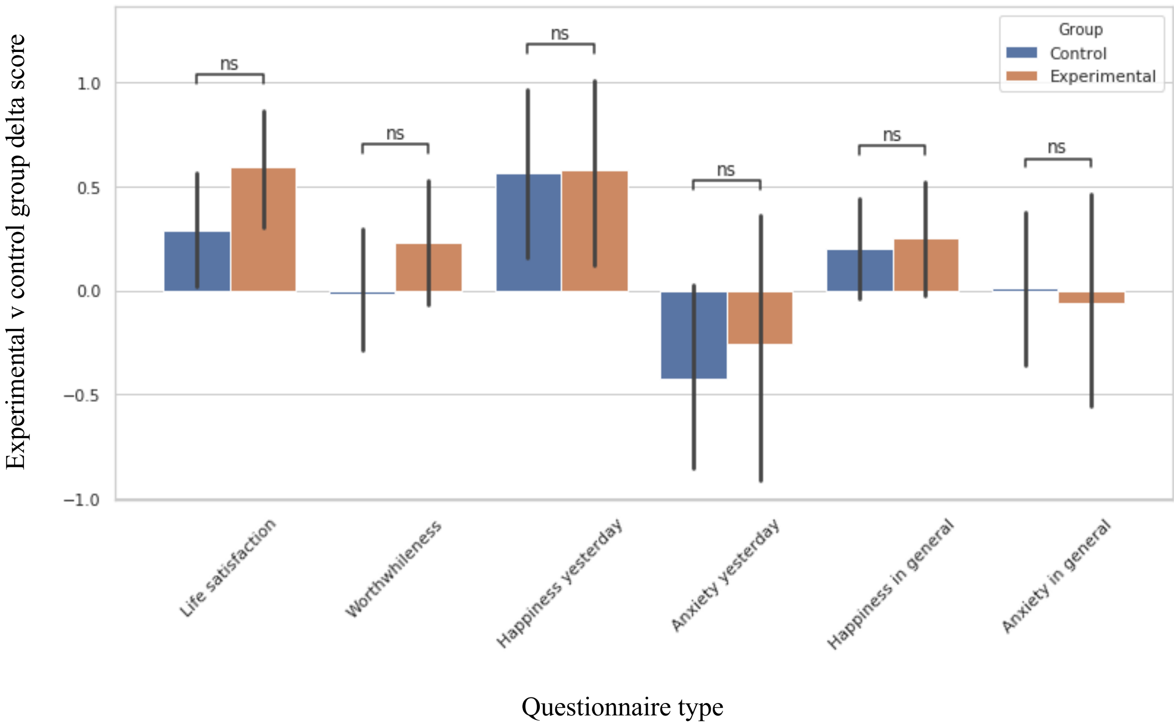


Figure 3. *Graph comparing questionnaire delta scores between experimental and control conditions for each outcome variable (week 4). Statistical significance was calculated using Mann Whitney U tests to compare delta (onboarding-exit) scores between experimental and control groups for each outcome variable. Two-sided p values reported: NS, not significant, *: p < 0.05, **: p < 0.01, ***: p < 0.001*

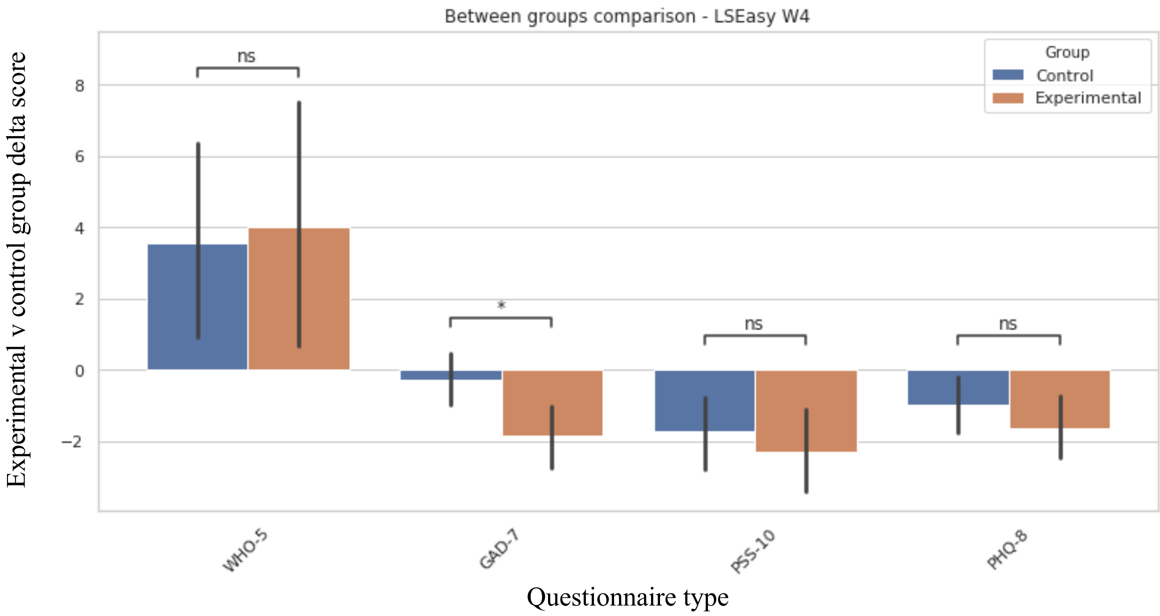


Figure 4. *Graph comparing questionnaire delta scores between experimental and control conditions for each outcome variable (week 4). Statistical significance was calculated using Mann Whitney U tests to compare delta (onboarding-exit) scores between experimental and control groups for each outcome variable. Two-sided p values reported: NS, not significant, *: p<0.05, **: p < 0.01, ***: p < 0.001.*

Our simple regression analyses also revealed that filling in more DRM questionnaires throughout the study was associated with a significantly larger reduction in clinical anxiety (b = -0.152, SE = 0.036, $p < 0.001$, $r^2 = 0.182$).

Discussion

Like studies 1 and 2, study 3 also revealed a positive impact of reporting WIM alongside contextual reports on anxiety. However, in this study, the effect was observed for clinical measures of anxiety (GAD-7). The between-group comparisons showed that the significant reduction in clinical anxiety in the treatment relative to the control group was large. Specifically, the effect was of similar magnitude to those identified in study 1.

Within-person comparisons revealed significant reductions in clinical anxiety within the treatment but not in the control condition, demonstrating an impact of WIM-reporting in the expected direction. Unlike studies 1 and 2, no significant effects were observed within the treatment group for anxiety yesterday or anxiety in general. This difference in impact might be since responses to outcome measures were collected at the 4-week mark, which was 2 weeks after participants stopped completing daily DRMs, rather than immediately after the end of the treatment, as in studies 1 and 2. It may also be because EMA was not used in addition to DRM.

Consistent with studies 1 and 2, no significant improvements in life satisfaction, worthwhileness, happiness yesterday or happiness in general were observed. Additionally, there were no significant improvements identified for stress, trait depression or evaluative subjective wellbeing as measured by the WHO-5. Importantly, this study has shown that a stand-alone DRM based intervention study (excluding EMA measures of WIM and associated context) can still be effective at reducing anxiety measured in clinical terms.

6. **General Discussion** (note this only relates to paper 4)

Heisenberg's landmark quote "to observe is to disturb" asserts that the simple act of observing alters the system that is studied (Heisenberg, 1930). When it comes to subjective wellbeing research, surprisingly little is known about how measuring people's subjective wellbeing impacts *the system* that is studied. The 3 studies conducted for this research assess the impact of reporting WIM and its associated context (thoughts, activities, location and company) using EMA and DRM questionnaires over the course of a few weeks on a range of mental wellbeing outcomes in a non-clinical population. Overall, we find that completing EMA and/or DRM questionnaires is related to significant reductions in anxiety across either single-item momentary measures or multiple-item clinical trait measures, in all studies.

In studies 1 and 2, we find medium-large effects of completing EMA and DRM questionnaires on reports of anxiety *yesterday* (only study 1) and anxiety *in general* (both studies) after three and two weeks, respectively. In study 3, we find a large effect of filling in DRM questionnaires for two weeks on clinical anxiety (i.e., GAD-7 scores) reported two weeks after the end of the study (4 weeks since the initial intervention began). The effect sizes we observe across studies (*d* = [0.18, 0.36]) are comparable to the effects of well established and refined behavioural interventions (*d* = [0.22, 0.44], Weiss et al., 2016) and positive psychology interventions on subjective wellbeing (*d* = [0.08, 0.43], Koydemir et al., 2021). This is particularly striking given that these measures were not initially designed to be interventions and only target one active ingredient: WIM-reporting. Reporting on WIM and its associated context did not impact happiness, worthwhileness or life satisfaction, though it did trend towards improvements on all mental wellbeing indices, with the exception of worthwhileness. We found no evidence that reporting WIM and its associated context improves stress, trait depression or evaluative wellbeing as measured by the WHO-5, though again these outcomes did trend towards improvements.

This work presents repeated evidence of a positive impact of reporting WIM and its associated context on anxiety. Studies 1 and 2 used an active control group, whereby participants in the control group received a sham version of the EMA and DRM questionnaires, whilst a passive (no intervention) control group was used in study 3. The sham version asked participants to report what they were doing, who they were with and where they were, but *not* what they were thinking about or how they were feeling. The fact that a positive impact of reporting WIM alongside context on anxiety remained in the study that used an active control suggests that the identified effects on anxiety are not being brought about by a placebo effect facilitated by the idea that an intervention is taking place and therefore likely to have an impact. The fact that reporting WIM and its associated context has an impact on anxiety over and above reporting about contextual factors alone also highlights the importance of combining WIM reports with contextual reports to generate anxiety reductions.

The studies presented in this paper find beneficial effects of using EMA and/or DRM questionnaires on people's reported levels of anxiety across both clinical (GAD-7) and population survey (ONS-4) measures of anxiety, different timespans, and, most importantly, different social and cultural contexts. The impact of answering EMA and DRM questionnaires on anxiety was captured first in a sample of working adults living in Spain and Spanish-speaking Latin American countries, then in a student sample in a UK university, suggesting that our findings may be generalisable across cultural contexts and occupations. Whilst we do note inconsistencies in terms of which anxiety measures were impacted across which study, there is a general pattern of improvements in anxiety across the board. We encourage future work to explore these inconsistencies in greater detail by replicating this work in clinical versus non-clinical samples of comparable status (e.g., students). These outcome differences also speak to the potential of this intervention to target multiple components of anxiety.

In addition to differences in average effects, we also found that completing more EMA and/or DRM questionnaires was consistently associated with greater reductions in the anxiety measures, suggesting that the impact of the intervention increases with more exposure, i.e., more reports. Furthermore, stronger anxiety reductions were found after three weeks in study 1 than after two weeks in study 2, suggesting also that the impact of WIM-reporting, especially on momentary measures of anxiety (i.e., anxiety yesterday), might develop over more extended periods of time. This is corroborated by the findings in study 3, where the impact on people's GAD-7 reports was found two weeks *after* the end of the study. Previous studies looking at the impacts of other wellbeing interventions have shown similar patterns of lagged improvements in clinical measures of mental health, suggesting that similar processes may be at play here (Catuara-Solarz et al., 2022).

It is interesting that reporting on WIM and its context leads to significant reductions in anxiety but not increases in happiness, worthwhileness, or life satisfaction. While it remains unclear why this type of intervention would reduce negative but not increase positive affect, this differential result is consistent with the large body of evidence supporting the independence of positive and negative affect (Diener & Emmons, 1984; Goldstein & Strube, 1994). Perhaps more puzzling is the fact that completing DRM in study 3 reduced anxiety according to GAD-7 scores, but not depression (PHQ-8), as other studies have hinted at. Since both constructs are closely tied to negative affect and have a strong association with each other, one might expect interventions reducing anxiety to also reduce depression.

There are, however, some important factors on which anxiety and depression differ. Namely, anxiety tends to be future-oriented and directed towards people's internal worlds, while depression tends to be past-oriented and related more to interpersonal aspects of reality (Pomerantz & Rose, 2018; Kreitler, 2018; Shipp & Aeon, 2019). Given that the process of WIM-reporting through EMA and DRM is inherently focused more on current (or recent) internal experiences, it may not be surprising that such reports would be more beneficial for dealing with anxiety than depression. It might also be that part of the benefit of these reports comes from people being able to project themselves into the near future with more clarity, therefore reducing future-related negative affect through better affective forecasting (Wilson & Gilbert, 2005). Since the present moment is often factored into our predictions about the future, orientating people's attention to current, or recent, thoughts and feelings may help to produce more optimistic affective forecasts, thereby reducing anxiety. Alternatively, it may be that rather than altering their affective response, people alter their behaviour. Though it should be noted that other studies exploring behavioural reactivity to EMAs have tended more generally to find null results (please refer to Introduction for an overview of these). While the present studies were not suited to test these potential mechanisms, future work could explore this by documenting people's motivations to change their affective and/or behavioural tendencies alongside frequent WIM-reporting and associated context.

Although some previous studies have found a direct impact of EMA reports on depression, the literature remains inconclusive as to whether EMA can be used as an intervention against depression. Studies using EMA as an accompaniment to other interventions (CBT and positive psychology) tend to report more consistent effects (Beams, 2021). Thus, although EMA shows promise in reducing depression, this may require more than mere reporting of WIM and its associated context. In line with this, depression treatment may require a more structured, person-facing, intervention, as well as deeper awareness levels. Nevertheless, given that anxiety commonly precedes depression (Cosci & Fava, 2021), the type of intervention we present in this study could be considered a useful preventative tool against the onset of depression. Moreover, the relative ease of implementing a stand-alone digital intervention like this is significant compared to more complex treatment methods like CBT (Wichers et al., 2011).

Limitations and Future Directions

There are some factors limiting the interpretation of our findings. Firstly, while our studies conclusively show that reporting on WIM and its associated context is an effective way to reduce anxiety, it is unclear which specific method (EMA or DRM) and which specific reports are the driving forces behind our findings. Interestingly it does appear that whilst EMAs used in combination with DRMs (as seen in studies 1 and 2) tended to impact anxiety yesterday and anxiety in general, we note that the DRM only study (3) impacted on the clinical anxiety measure only. Future work may benefit, therefore, from testing the differential impact of these measures when used as stand-alone interventions (e.g. conducting an EMA only study alongside a DRM only study). Additionally, since no studies in this paper explored the impact of wellbeing-specific EMA and DRMs (without context), it also remains unclear whether what matters is

WIM reports only, or WIM combined with contextual reports. Building on this, future research should explore the relative contribution of these factors when considered in isolation, and in tandem, with interventions that focus on different aspects of the reflection process, to see which processes are most effective at improving which facets of mental wellbeing.

It is also important to note that we collected data for study 3 during the COVID-19 pandemic, whereas study 1 and 2 were conducted before the pandemic. While it is unclear how this difference in context might have affected the way people reacted to completing DRM questionnaires, we know that the pandemic had a large impact on mental wellbeing in the UK and across the world (Helliwell et al., 2021). The unique social context of the pandemic may call into question how generalisable results from this time period are relative to “normal” times. As such, future research should consider replicating the findings of study 3 in a post-pandemic context.

Finally, the design of studies 1, 2 and 3 did not allow us to unpack the underpinning mechanisms that led to the observed reductions in anxiety. The act of reporting may draw people’s attention to their feelings and the context in which these are experienced, thereby improving their awareness of how they feel in certain situations, or it might improve mental wellbeing via improvements in emotion differentiation. Further research is needed to test these possibilities. Nevertheless, this work extends current literature on wellbeing interventions by showing that identifying and contextualising WIM holds strong potential as an overall mental wellbeing enhancing intervention.

Conclusions

Anxiety is the most prevalent mental health disorder in Europe today, with an estimated 25 million people suffering with its effects, whilst many more experience its effects across the globe (OECD, 2018). This work has shown that asking people to report their WIM and its associated contexts through EMA and DRM questionnaires, over periods of as little as two weeks, significantly reduces anxiety, and more frequent reports are associated with greater reductions in anxiety. Therefore, what have thus far mostly been used as data collection tools could turn out to be powerful, low-cost interventions with large scale impact following further replication and refinement. Indeed, reducing anxiety may be as simple as taking a few minutes every day to report what we did, thought about and how we felt in different situations.

Data availability

The data that support the findings of this study are available from Koa Health but GDPR restrictions apply to the availability of these data, and so they are not publicly available. Data are however available from the authors upon reasonable request and with permission of Koa Health.

Personal reflections on paper 4

These are personal research reflections and do not form part of the main paper. The purpose of this section is to highlight analysis challenges and key research skills developed.

Analysis challenges

In this analysis it was necessary to consider many different factors that might influence our results. For example, it was necessary to check not only if there were mean overall differences between the treatment and control groups but also what was driving those differences (e.g., increases in mental wellbeing following the treatment or reductions in mental wellbeing following the control). Moreover, if the intervention was successful, we might expect that higher dosage of the intervention (i.e., more frequent WIM-reporting) would lead to greater changes. This was explored with the regression analyses.

To better isolate the impact of the intervention it was also important to consider whether those that dropped out of the study might have differed from those that stayed in. For example, it might have been that this type of intervention only works for people that are high in mental wellbeing to begin with and those are the people that stayed in. Importantly, however, it was found that those who dropped out of the study were similar in mental wellbeing and demographic profile to those that remained. Another small but important detail that was checked was whether there were differences in WIM reports depending on what time of the day people responded. No significant differences were identified here.

Key research skills developed

I played a lead role in designing and collecting the data for the three RCTs involved in this paper. I also led the development of the research ready mobile application was used to collect SWB reports, coordinating insights from computer scientists and app developers to produce the final product.

I also played a lead role in orchestrating the collaboration between Alpha (the mental health intervention company that funded the studies) and the LSE. I was the main point of contact between these organisations, and I ensured that both parties were able to benefit from the research output. This involved making strategic decisions about how many, and which questions, to include. For example, I created a specialised drop-down list of activities for LSE students to reduce friction in the WIM-reporting phase. I also made sure that the EMA and DRM questions used in each questionnaire were reflective of previously validated research on each questionnaire.

During the process of conducting this research I became well acquainted with the relative advantages of using different types of control groups (active versus passive control). For example, I now understand that both options can be usefully combined to better assess the true impact of an intervention. Whilst an active control group can function in a way that is similar to a placebo, by administering the intervention without the active ingredient (thereby demonstrating impact over and above treatment expectation) passive control can reveal important detail about the counter factual for individuals undergoing no perceived treatment at all.

4. Critical discussion

Mental wellbeing continues to be listed as a top priority for most people across the world. The social, health and economic costs of not providing for mental wellbeing are profound (Layard, 2018). Growing expenditure on mental healthcare poses a serious challenge for governments worldwide (Patel et al., 2018). Wellbeing-in-the-moment is an important determinant and consequence of mental wellbeing and therefore presents exciting potential to help advance our understanding of it. This thesis has sought to harness the power of WIM to 1) improve our understanding of how to measure the core components of mental wellbeing: subjective wellbeing (SWB) and mental illness, and 2) identify important WIM-related pathways that can be used to influence mental wellbeing.

The following two Sections, 4.1. “Measure: overarching contributions” and 4.2. “Intervene: overarching contributions”, discuss the overarching contributions of papers 1-2 (measure) and 3-4 (intervene) to these aims. Whilst the individual contributions of each paper can be found in Chapter 3 “Empirical Work” this chapter focuses on how the *combined* papers from each Measure and Intervene section contribute to measurement and intervention-based knowledge. This chapter does not cover limitations since these are specific to each study and can therefore be found in their respective papers in Chapter 3. The final Section 4.3. “Introducing the WIM Intervention Framework” brings together the insights from both sections of the thesis and considers these in the development of a novel framework. The WIM Intervention Framework offers a foundation for continued investigation into WIM-related pathways to mental wellbeing that aims to facilitate mental wellbeing conceptualisation and intervention.

4.1. Measure: overarching contributions of papers 1 and 2

Before exploring WIM-related pathways as a novel means to intervene on overall mental wellbeing, a clear understanding of how best to measure the core constructs that make up overall mental wellbeing – SWB and mental illness – and their relationship to WIM, is necessary. Papers 1 and 2 have contributed to these aims.

Paper 1 has done so by comparing the two most popular WIM-based SWB measures: the Ecological Momentary Assessment and the Day Reconstruction Method. The key findings of this paper were that 1) the duration of WIM experiences does not show up in experiential SWB reports using either of these popular measures and 2) the measures produce comparable SWB scores. Paper 2 has done so by conducting a two-study replication in which popular anxiety and depression questionnaires were filled out following different WIM inductions (happy/sad/relaxed/anxious/neutral). The key findings of this paper were that anxiety questionnaires are largely unchanged by WIM inductions, whereas small but significant impacts of WIM inductions were identified for depression questionnaires.

Together these findings have raised some noteworthy concerns relating to the validity of SWB and mental illness questionnaires. With respect to paper 1, it is a substantive fact that feeling happy for longer will make you happier overall and so experiential SWB measures ought to have picked up on this. Therefore, this finding calls into question the ability of existing WIM-based experiential SWB measures to capture the duration of affective experiences. As such researchers interested in the relationship between duration and experiential SWB are advised to invest some time and effort into the development of new and better ways of capturing affective duration. With respect to paper 2, this work provides empirical support for the assumption that depression, as it is currently measured, contains at least some aspects that are influenced by, and overlap with, WIM.

Additionally, both papers have contributed to developments in our understanding of the interrelationship between WIM, SWB and mental illness. Paper 1 has done so by demonstrating that both momentary and WIM-based experiential SWB reports (as per the Ecological Momentary Assessment measure) and recently recalled WIM-based experiential SWB reports (as per the Day Reconstruction Method which relies on WIM-reports from the previous day) produce equivalent SWB reports. As such, there is no need to differentiate between WIM experiences in the moment and WIM recollections of the previous day when approximating SWB.

Paper 2 has done so by highlighting a small degree of overlap between WIM and depression. Specifically, it was found that increases in arousal (activated/alertness) were associated with decreases in trait depression reports. In addition, increases in positive valence were associated with decreases in state depression. If an impact of WIM on mental illness continues to be replicated and affirmed by future studies, it will be up to researchers to decide how best to integrate these insights. For instance, they may want to keep mental illness measures as they are and allow for this overlap since it is small, and consistent with disease models of anxiety and depression, that conceptualise these illnesses as lying on a continuum with WIM (Ford et al., 2015). Alternatively, if greater construct discrimination is required, which can enable a better understanding of how different mental illnesses relate to WIM in isolation, then these measures may need to be reconsidered to ensure that WIM-based components are removed. Given that the present work is limited in terms of WIM-induction methods tested, it is vital that future work trials different variations of WIM-induction, such as film clips, music and naturally occurring mood changes, before these conclusions are taken as confirmatory.

The schematic diagram in Figure 1 presents these insights by showing the extent of identified empirical overlap between WIM, SWB and mental illness. Here, we can see that WIM and SWB overlap insofar as experiential SWB is concerned. This is because WIM feeds directly into the measurement of overall experiential SWB and should therefore be considered as an inherent part of this construct. Evaluative SWB, however, forms the non-overlapping part of SWB that is generally unrelated to WIM, according to the most recent and empirically robust work reviewing this relationship (Yap et al., 2017).

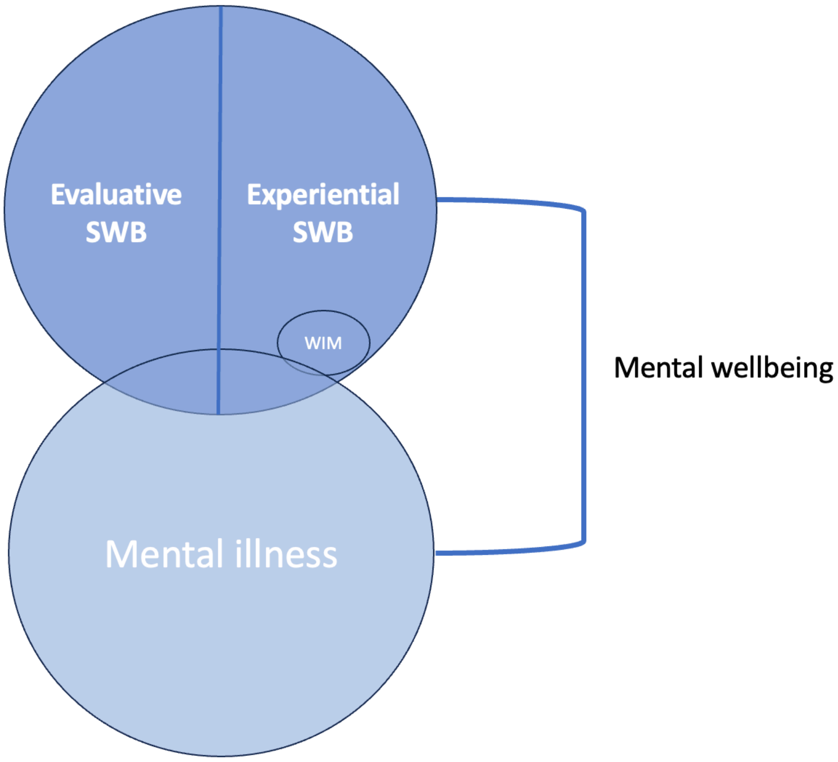


Figure 1. *A schematic diagram visualising the extent of overlap between WIM, SWB and Mental Illness (anxiety and depression) highlighted by empirical work exploring the interrelationship between these constructs.*

Mental illness, by contrast, is mostly distinct from WIM. It does, however, still overlap to a small but significant degree as highlighted in paper 2. In general, the more studies conducted that can help to affirm, or disaffirm, the relative positions in the diagram above the more accuracy we can attribute to our current understanding of these constructs. Psychological findings may be strengthened and enriched by the incorporation of neuroscientific findings to assist this conceptualisation. For example, studies exploring differences in brain activation patterns between WIM, SWB and mental illness may help to identify additional points of overlap or distinction between these constructs that are not easily detected by self-report measures. Greater empirical support for conceptual distinctions can prove vital in helping to promote agreement amongst psychologists as to exactly what constitutes these constructs. Since intervention assessment is dependent on clear understanding of how to conceptualise and measure desired outcomes, alignment of this kind can lead to more effective and streamlined theory and intervention (Thornicroft & Slade, 2014).

Overall, papers 1 and 2 have developed our understanding of the relationship between WIM with SWB and mental illness providing important methodological insights that raise questions about the measurement validity of experiential SWB and depression and contribute to our understanding of the interrelationship between WIM and overall mental wellbeing. The findings are also useful in informing the final two papers of this thesis, which focus on how to intervene on overall mental wellbeing. The experiential SWB measurement insights obtained from paper 1 were utilised in paper 3, which used the same experiential SWB measures to assess whether using these measures regularly, impacts on overall mental wellbeing. The measurement insights obtained from paper 2 were utilised in papers 3 and 4, since in assessing the impact of any WIM-related behaviour on mental illness, it is useful to know the degree of initial overlap between these constructs. For instance, knowing that WIM and mental illness share at least a small degree of WIM related variance, tells us that similar processes may affect both constructs, therefore WIM interventions might also impact mental illness and vice versa.

4.2. Intervene: overarching contributions of papers 3 and 4

An important first step towards designing effective mental wellbeing interventions is identifying and understanding the pathways that lead to improvements in mental wellbeing outcomes. Papers 3 and 4 have both contributed to this overarching aim by highlighting WIM-related behaviours that influence mental wellbeing, or mental wellbeing associated outcomes.

Paper 3 has done so by assessing whether WIM-related social behaviours taking place in a natural setting are associated with mental wellbeing related approach-avoidance tendencies. The key findings of this paper are that naturally occurring reductions in overall interaction (digital and in-person) were significantly associated with a decreased tendency to approach relative to avoid sad faces (i.e., decreased sad tendency). In other words, people became more avoidant than approaching of sad faces over time in response to reduced overall interaction. A separate association was also found between generalised anxiety and stress with sad tendency, whereby increases in anxiety and stress also related to decreases in people's tendency to approach relative to avoid sad faces over time.

Paper 4 has done so by conducting three RCTs exploring whether frequently reporting WIM and its associated context improves mental wellbeing outcomes over time. This body of work identified a persistent and significant negative impact of WIM reporting on anxiety when comparing pre and post measures. Importantly, however, the type of anxiety in which reductions were observed (e.g., state/trait) differed across studies. Reporting on WIM did not impact overall SWB measures of happiness, worthwhileness, or life satisfaction however, suggesting a direct and independent pathway from WIM to anxiety.

Together these findings have succeeded in highlighting two novel WIM-related pathways to improved mental wellbeing. With respect to paper 3, it appears that reducing social behaviour results in an increase in automatic avoidance of sad faces, which may function to bolster against the negative impact of reduced social behaviour on mental wellbeing in the longer term. As such, this work has demonstrated potential for approach-avoidance as a potential protective mechanism that may shield individuals from affective decline. This finding is novel; whilst it has been shown that altering approach-avoidance tendencies can change behaviour, to my knowledge, this relationship has never been explored in reverse. This research has also affirmed the strong association between mental wellbeing (anxiety) with approach-avoidance in a naturally occurring setting and highlighted a lesser-known association between WIM (stress) and approach-avoidance.

With respect to paper 4, highlighting a direct route from WIM-reporting to anxiety reduction, has identified WIM-reporting as a successful pathway to improving mental wellbeing. The size of the identified effect is comparable to that obtained by other leading behavioural interventions (e.g. wellbeing therapy, Fava et al., 2005; and acceptance and commitment therapy, Hayes et al., 2013) and positive psychological interventions (Lee Duckworth et al., 2005). However, WIM-reporting is a less onerous process that is not dependent on specific medical training or support making it easier to administer at scale. Importantly, this effect was consistent across different social and cultural contexts (study 1: UK adults and adults living in Spain and Spanish speaking Latin-American countries, study 2 and 3: UK students). These insights highlight potential merit in the development of novel WIM-reporting interventions that can be used to reduce anxiety.

Overall studies 3 and 4 have contributed to the mental wellbeing intervention literature by identifying important WIM-related pathways to mental wellbeing. These findings have the potential to inform powerful, low-cost interventions aimed at improving mental wellbeing.

4.3. The WIM Intervention Framework

Together, the studies that make up this thesis have highlighted the potential of WIM to reveal new insights about how best to measure and influence overall mental wellbeing. These findings are useful in isolation; however, their impact can be better understood once considered within the broader context within which they belong. One way to assist the contextualisation of WIM-based research findings is to generate a framework that enables WIM-based intervention components to be mapped onto mental wellbeing outcomes. Since mental wellbeing interventions often involve a complex combination of different pathways to overall mental wellbeing (Moore et al., 2015), isolating and differentiating between the components that affect change is an important prerequisite to understanding why a given intervention might work and when it may be most effective (Wight et al., 2016). This section presents a recommended framework, the WIM intervention framework (WIMIF), that facilitates such a mapping.

The WIMIF takes inspiration from Dynamic Systems Theory, which contends that multiple forces interact to determine a change in any system (Poincaré, 2017). Within dynamic systems, all included factors are important in determining an outcome (Perone & Simmering, 2017). For example, the act of turning on a light switch may be explained by several mental, physiological, and motivational orientations that come together to determine that action. If any of these components is taken away the end behaviour may be subject to change. Whilst computer science disciplines have demonstrated the merit of network analysis in helping to elucidate unknown pathways to mental wellbeing intervention, this is an inductive approach. Such research can be complemented and enhanced by an adjacent stream of intervention research that highlights network associations within a pre-considered theory-driven framework. Combining inductive and deductive research methods can help to provide a more complete understanding of a given area (Bonner et al., 2021).

In general, there is a lack of systematic guidance on how to integrate theory-driven research on potential pathways to intervention (Van Valkengoed et al., 2022). Whilst independent models exist for cognitive (Young et al., 2014) and behavioural interventions (Walker, 2017) as well as in combination (Van Blisen, 2018) there have been no systematic efforts to demonstrate how a given set of WIM-related intervention pathways may be used to improve mental wellbeing. As such there is little understanding about how and whether existing interventions function through, or via the means of, WIM change. Moreover, despite an increasing move towards spectrum-based psychotherapy models which conceptualise WIM and overall mental wellbeing as existing on a spectrum with one another, empirical and theoretical work that explores how and when transition occurs along this spectrum is lacking (van Agteren, 2021).

Within this context, the WIMIF framework aims to provide three key benefits to mental wellbeing researchers and practitioners. First, by visualising different pathways from WIM to overall mental wellbeing within a broader structure, it highlights important commonalities and differences between how overall mental wellbeing outcomes relate to WIM-related components. These insights can be used to improve the conceptualisation of WIM and mental wellbeing related constructs.

Second, visualisation of multiple associations within one common structure can reveal how different associations might be usefully *combined* to determine an effect. For example, once we know which WIM-related factors promote *both* SWB and mental illness, individuals requiring improvements in both outcomes can be given interventions containing those factors, leading to faster and broader psychological improvements. Similarly, an individual low on one dimension (e.g. SWB only) can be presented with a more targeted and streamlined intervention including components that specifically target that outcome. Moreover, by visualising different pathways in one common framework, different means of achieving the same outcome can be directly compared to each other.

Third and finally, by visualising how research findings can inform and build upon each other, the WIMIF provides a motivation for researchers to work together and replicate results in this area, enabling higher levels of understanding and methodological rigour (Walker et al., 2014).

Figure 2 below depicts the proposed framework. Level 1 at the bottom of the figure contains any WIM experience, including discrete affective experiences (e.g., happiness, sadness and so on) or overarching affective components such as valence and arousal. Level 2 consists of processes and behaviours that are associated with both WIM and overall mental wellbeing. Any isolated process or behaviour that has a strong, direct association with both WIM and overall mental wellbeing outcomes can be considered for this level. For example, social interaction can be considered for this level since people typically experience negative affect when opportunities to interact are removed (Cacioppo & Patrick, 2008). If prolonged, reducing social behaviours can also impact SWB (Sun et al., 2020) and mental illness (Yanos et al., 2001). In the present framework, processes/behaviours connecting WIM to mental wellbeing are considered important in addition to WIM only, since many interventions do not affect change directly and instead work through targeting associated processes/behaviours (Michie et al., 2018). Level 3 contains mental wellbeing outcomes including SWB and mental illness (anxiety and depression).

Within this framework, it is important to stress that whilst the transition between the lower levels to the higher levels is most interesting for understanding how to impact overall mental wellbeing, the relationships between the different entities can work in either direction. Note that it is not necessary for all levels to be activated at the same time to determine a change in overall mental wellbeing outcomes. In helping to visualise when and under which conditions different levels activate in unison, a deeper understanding of how to facilitate overall mental wellbeing outcomes can be developed through this framework. For example, if researchers input key information about the measures and contexts used when entering study results, it should be possible to explore commonalities between these for circumstances where entities across multiple levels are activated together.

It is recommended that insights from studies about the impact of different intervention components linking WIM to overall mental wellbeing are added to the framework after those studies have been completed so that results from different studies may be combined. In this way, over time, it will be possible to build a more complex system of associations to better aid our understanding of how transitions are made between WIM and overall mental wellbeing. The thickness of the lines linking each factor can be increased or decreased depending on the strength of the evidence found to be associated with that link over time. Research quality may be scored on factors like causality, replication, sample size, sample representation, external validity, and number of null findings. Arrows represent causal relationships and lines represent non-causal relationships.

The WIM Intervention framework

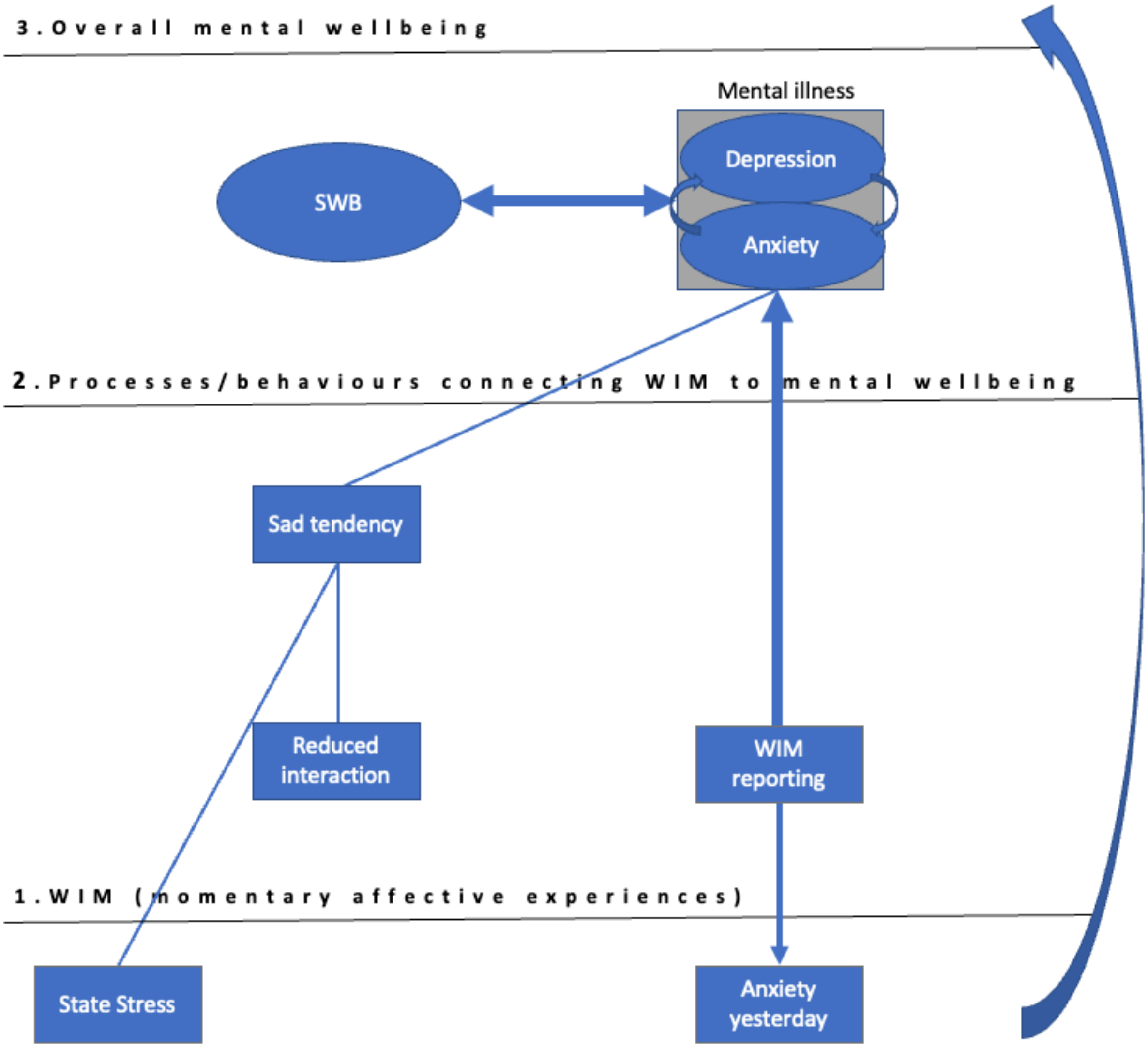


Figure 2. *A visualisation of the WIM intervention framework with results inserted from papers 3 and 4. Single lines between boxes represent direct associations between different entities that can be either positive or negative. The line thickness can be increased or decreased depending on the strength of the research supporting that association (greater thickness represents higher quality empirical support).*

To highlight the benefits of being able to situate the impact of key WIM, SWB and mental illness related processes and behaviours within a broader network of similar affective constructs and associations, the results from intervention papers 3 and 4 have been incorporated into the framework above. Here it can be seen that both studies have uncovered potential WIM-related pathways to reducing trait anxiety. Based on the strength of the evidence obtained in these studies, WIM-reporting appears to be the most promising route. The pathway from WIM-reporting to trait anxiety reduction (demonstrating the findings observed in paper 4) is thicker than the pathway from sad tendency to anxiety due to the higher strength of this research: similar observations were discovered in three causal studies. The arrows represent the direction of findings observed. For example, since a causal impact of WIM-reporting was identified for both trait anxiety (level 3) and anxiety yesterday (level 1) there are arrows leading from WIM-reporting out to these constructs. It is important that both impacts on mental illness and WIM measures are visualised here since this demonstrates potential for a positive feedback loop whereby the intervention impacts trait anxiety symptoms directly as well as via reductions in momentary anxiety, which may feed back into trait anxiety reductions over time. If true, this could lead to even greater impact and faster transition from mental illness to healthy WIM. Understanding whether interventions that impact WIM and mental illness directly prove more efficacious than those that do not is another interesting insight that this framework facilitates.

The line representing a pathway from social interaction to anxiety (demonstrating the findings observed in paper 3) is thinner and arrowless since, unlike paper 4, these findings were not causal and were reliant on the results of just one study conducted in a pandemic context. This figure shows that reduced interaction, state stress and anxiety are all associated with sad tendency. By visualising pathways in this way new suggestions for combination interventions become apparent. For instance, combining social interaction with stress reducing interventions may be a particularly effective means by which to improve overall mental wellbeing, with a specific potential benefit for those suffering with anxiety. Of course, this picture may change once the findings of further research have been incorporated in the visualisation.

Since it is highly challenging, if not impossible, to measure and account for every single component in a complex system that determines a change in overall mental wellbeing, the present classification is limited to WIM-related components. A focus on exploring WIM-related intervention components in isolation is important since it may reveal insights relevant to developing our understanding of how WIM relates to overall mental wellbeing that are otherwise missed by more general models. For example, The Generic Cognitive Model (a recent adaptation of Beck's Cognitive Therapy model) is useful for understanding how common cognitive processes contribute to specific disorders (Beck, 2014). The main tenet of this theory is that cognitive schemas (specific ways of thinking) are responsible for the maintenance of psychological disorders. These schemas are thought to be influenced by a series of factors including genetics, physiology, attention, and memory, all of which can change or be changed by environmental triggers. Schema distortion is proposed as the core means by which psychopathology is determined.

Crucially however, in its emphasis on cognitive factors this framework is missing important information about how mental illness relates to subjective wellbeing and WIM. For example, it proposes that distorted schemas trigger maladaptive affective, behavioural, and motivational responses (Beck, 2014). However, it does not consider that affective experiences might also feed into those schemas, or other processes that separately determine mental illness (e.g., Croker et al., 2013). In one past example of this, a drop in negative thought patterns was identified following remission from an episode of depression (Simons et al., 1984), suggesting that affective experience might also determine thought patterns. This possibility can only be explored when one considers WIM as an input as well as an outcome to mental illness, and overall mental wellbeing more generally.

A stream of recent work in psychotherapy has sought to develop a similar mapping of symptoms onto mental illness diagnoses (McNally, 2021). This line of work draws upon Network Analysis (NA) to help identify patterns of associations that are common for different mental illnesses. NA uses a series of techniques to detect associations between different "nodes" (objects of study). It uses these associations to better understand outcomes (e.g., mental illnesses) that arise from the recurrence of these associations. The basic premise is again, that change can be better accounted for when we understand interrelations among the entities that cause it. A similar line of analysis may be taken to mathematically approximate the relative importance of different WIM-related pathways to overall mental wellbeing. This kind of analysis must be adopted with caution, however, since results can be unreliable and there is still some work to be done to ensure that conclusions are clinically valid (Contreras et al., 2019). It may prove advantageous for computer scientists to develop a simplified version of this whereby the input parameters are determined by the researcher for now, taking on board the criteria I have specified above with respect to research quality. The ability to filter the visualisation by different factors (e.g., natural experiments or certain constructs of interest) will also be advantageous.

As well as zooming out by combining WIM-related pathways from different studies, it will also be necessary to zoom in and conduct more focused research on the pathways that over time reveal themselves as having the greatest potential to affect overall mental wellbeing improvements at the broadest level. When zooming in, it may be particularly beneficial for researchers to focus on testing subtle adjustments to the intervention components highlighted as having the most potential. One new and potentially impactful way to do this is by employing a recently developed "megastudy" paradigm. The megastudy is a large-scale field experiment whereby researchers come together to test multiple intervention variations targeting the same pathway in the same context. This allows for a fair evaluation and comparison of different intervention implementations within the same experimental context (Milkman et al., 2021). Such studies can help to avoid the risks of comparing 'apples and oranges' that may come from focusing only on associations at the broader level. They can also be used to explore individual variation in treatment effects, that can be used to inform intervention personalisation (Gan et al., 2022; Whiston et al., 2019).

In summary, the WIMIF has a series of key strengths that can enable an acceleration of rigorous research into WIM-related pathways to overall mental wellbeing. It provides a means to visualise different pathways to overall mental wellbeing, situating the impact of key WIM, SWB and mental illness related processes and behaviours within a broader network of similar affective constructs and associations. This integration of SWB and mental illness outcomes, as well as the identification of associated pathways to altering them, is crucial for identifying shared and distinct features of the core mental wellbeing constructs that can be used to design more targeted and effective interventions. The WIMIF also highlights areas where intervention components can be strategically combined to achieve an impact on overall mental wellbeing. Finally, it provides a motivation for research collaboration and replication by highlighting how different findings (including null findings) can contribute towards a broader understanding of the interrelationship between WIM and overall mental wellbeing.

Overall conclusion

In conclusion, this thesis contributes to the growing body of research on the interrelationship between WIM and overall mental wellbeing by exploring the complex and dynamic interrelationship between WIM, SWB and mental illness. It has made important contributions informing how we measure and influence overall mental wellbeing. In terms of measurement, it has probed the relative utility and validity of two popular WIM-based experiential SWB measures. It has also highlighted small but significant overlaps between WIM and popular depression measures which has provided some important clarity on the shared components that WIM and mental illness measures detect. In terms of intervention, it has identified a significant association between WIM-related behaviour (reduced social interaction) and mental wellbeing-related determinants (approach-avoidance), highlighting these components as potential pathways from WIM to mental wellbeing. By demonstrating a positive impact of WIM reporting on mental illness via reduced anxiety, it has also shown that WIM reporting has promise as a simple yet highly effective mental illness intervention.

Finally, whilst WIM is frequently conceptualised as existing on a spectrum with mental illness, we know little about how the transfer from WIM to mental illness occurs and which factors promote and inhibit these transfers. To assist the development of this line of enquiry and inform intervention development this thesis concluded by presenting a novel framework that can be used by researchers and practitioners seeking to better understand the interrelationship between WIM, SWB and mental illness, and identify novel WIM-related paths to mental wellbeing intervention.

Taken together, these insights have contributed to a meaningful line of enquiry into WIM-based opportunities for improving overall mental wellbeing measurement and intervention. In conclusion, it would appear that WIM is indeed more than just a feeling, it is also a promising pathway to improved mental wellbeing.

5. Introduction references

Aafjes-van Doorn, K., & Barber, J. (2017). Systematic Review of In-Session Affect Experience

Adler, M. D., Dolan, P., Henwood, A., & Kavetsos, G. (2022). “Better the devil you know”: Are stated preferences over health and happiness determined by how healthy and happy people are?. *Social Science & Medicine*, *303*, 115015.

Amini, S., Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2012). Fixed vs random: The Hausman test four decades later. In *Essays in honor of Jerry Hausman* (Vol. 29, pp. 479-513). Emerald Group Publishing Limited.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, ***15*****(6),** 713-724 (2001).

Baston, C. D., Shaw, L, L., Oleson, K. C. (1992). Differentiating Affect. Mood, and Emotion: Toward Functionally based Conceptual Distinctions (chapter). In Clark, Margaret S. (Ed), (1992). Emotion. Review of Personality and Social Psychology. Thousand Oaks, CA, US: Sage Publications.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry*, *4*(6), 561-571.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*, *11*(2), e0148037.

Bower, G.H. (1981). Mood and memory. Am. Psychol. 36, 129–148.

Briki, W. (2018). Trait self-control: Why people with a higher approach (avoidance) temperament can experience higher (lower) subjective wellbeing. *Personality and Individual Differences*, *120*, 112-117.

Butler, T., Pan, H., Tuescher, O., Engelien, A., Goldstein, M., Epstein, J., ... & Silbersweig, D. A. (2007). Human fear-related motor neurocircuitry. *Neuroscience*, *150*(1), 1-7.

Cabanac, M. (2002). What is emotion?. *Behavioural processes*, *60*(2), 69-83.

Clark, A. (2005). *Causes, Role, and Influence of Mood States*. New York: Nova Biomedical

Chalder, M., Wiles, N. J., Campbell, J., Hollinghurst, S. P., Haase, A. M., Taylor, A. H., ... & Lewis, G. (2012). Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *Bmj*, *344*.

Chamberlain, K. (1988). On the structure of subjective well-being. *Social indicators research*, *20*(6), 581-604.

Csikszentmihalyi, M., & Larson, R. (2014). *Flow and the foundations of positive psychology* (Vol. 10, pp. 978-94). Dordrecht: Springer.

De Neve, J. E., Clark, A. E., Krekel, C., Layard, R., & O'donnell, G. (2020). Taking a wellbeing years approach to policy choice. *bmj*, *371*

De Lange*, A. H., Taris, T. W., Kompier, M. A., Houtman, I. L., & Bongers, P. M. (2004). The relationships between work characteristics and mental health: Examining normal, reversed and reciprocal relationships in a 4-wave study. *Work & Stress*, *18*(2), 149-166.

Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J., & Steptoe, A. (2010). A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Social Indicators Research*, *99*(2), 269-283.

Dolan, P. (2015). *Happiness by design*. London: Penguin.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology*, *29*(1), 94-122.

Dolan, P., & Kudrna, L. (2016). Sentimental hedonism: pleasure, purpose, and public policy. *Handbook of eudaimonic well-being*, 437-452.

Dolan, P., Kudrna, L., & Stone, A. (2017). The measure matters: An investigation of evaluative and experience-based measures of wellbeing in time use data. *Social Indicators Research*, *134*(1), 57-73.

Dolan, P., Layard, R., & Metcalfe, R. (2011). Measuring subjective well-being for public policy.

Dolan, P., & Powdthavee, N. (2012). Thinking about it: A note on attention and well-being losses from unemployment. *Applied Economics Letters*, *19*(4), 325-328.

Dolan, P., & Metcalfe, R. (2010). 'Oops… I did it again': Repeated focusing effects in reports of happiness. *Journal of Economic Psychology*, *31*(4), 732-737.

Dormann, C., & Zapf, D. (2002). Social stressors at work, irritation, and depressive symptoms: Accounting for unmeasured third variables in a multi-wave study. *Journal of Occupational and Organizational Psychology*, *75*(1), 33–58. doi:10.1348/096317902167630

Ekman, P. Universals and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), Nebraska Symposium on Motivation (Vol. 19). Lincoln: Universiry of Nebraska Press, 1972.

Eysenck, M. W., & Fajkowska, M. (2018). Anxiety and depression: toward overlapping and distinctive features. *Cognition and Emotion*, *32*(7), 1391-1400.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370-378.

Freud, S. (1899/1999). The interpretation of dreams, trans. J. Crick. Oxford University Press. (Original work published in 1899.)

Gaudry, E., Vagg, P., & Spielberger, C. D. (1975). Validation of the state-trait distinction in anxiety research. *Multivariate behavioral research*, *10*(3), 331-341.

Göritz, A. S. (2007). The induction of mood via the WWW. *Motivation and Emotion*, *31*(1), 35-47.

Gross, J. J. (2010). The future's so bright, I gotta wear shades. *Emotion review*, *2*(3), 212-216.

Griffin, J. (1986). *Well-being: Its meaning, measurement and moral importance*. Clarendon press.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, *42*(4), 1009-1055.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annu. Rev. Clin. Psychol.*, *3*, 1-27.

Krekel, C., De Neve, J. E., Fancourt, D., & Layard, R. (2021). A local community course that raises wellbeing and pro-sociality: Evidence from a randomised controlled trial. *Journal of Economic Behavior & Organization*, *188*, 322-336.

Layard, R., & Clark, D. (2015). *Thrive*. London: Penguin books.

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., &

Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz social anxiety scale. *Psychological medicine*, *29*(1), 199-212.

Hertel, G., Neuhof, J., Theuer, T., & Kerr, N. L. (2000). Mood effects on cooperation in small groups: Does positive mood simply lead to more cooperation?. *Cognition & emotion*, *14*(4), 441-472.

Hodges, L. J., Walker, J., Kleiboer, A. M., Ramirez, A. J., Richardson, A., Velikova, G., & Sharpe, M. (2011). What is a psychological intervention? A metareview and practical proposal. *Psycho-oncology*, *20*(5), 470-478.

Izard, C. (1977). *Human Emotions*. New York: Plenum Press.

Izard, C. (1972). Patterns of emotions. New York: Academic Press, 1972.

James, W. (1884). What is an emotion?

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, *146*(4), 355.

Iasiello, M., & Van Agteren, J. (2020). Mental health and/or mental illness: A scoping review of the evidence and implications of the dual-continua model of mental health. *Evidence Base: A journal of evidence reviews in key policy areas*, (1), 1-45.

Kazdin, A. E. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy research*, *19*(4-5), 418-428.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776-1780.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *science*, *312*(5782), 1908-1910.

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, *24*(1), 10-16.

Kelloway, E. K., & Francis, L. (2013). Longitudinal research and data analysis.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of consulting and clinical psychology*, *73*(3), 539.

Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *American psychologist*, *62*(2), 95.

Keyes, C. L. (2013). Promoting and protecting positive mental health: Early and often throughout the lifespan. In *Mental well-being* (pp. 3-28). Springer, Dordrecht.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932-932.

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological science*, *23*(10), 1086-1091.

Knabe, A., Rätzel, S., Schöb, R., & Weimann, J. (2010). Dissatisfied with life but having a good day: time-use and well-being of the unemployed. *The Economic Journal*, *120*(547), 867-889.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509-515.

Kron, A., Pilkiw, M., Banaei, J., Goldstein, A., & Anderson, A. K. (2015). Are valence and arousal separable in emotional experience? *Emotion*, *15*(1), 35.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*, *38*.

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. *Psychological bulletin*, *137*(5), 834.

Lewis, M., Haviland-Jones, J., & Barrett, L. (2008). *Handbook of Emotions*. New York: Guilford Press.

LSE News (2022, March 3). *Mental health problems cost UK economy at least £118 billion a year – new research.* Retrieved from https://www.lse.ac.uk/News/Latest-news-from-LSE/2022/c-Mar-22/Mental-health-problems-cost-UK-economy-at-least-118-billion-a-year-new-research

Martin, P. P. (2022). Neural systems underlying approach and avoidance in anxiety disorders. *Dialogues in clinical neuroscience*.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology*, *67*(3), 267-277.

McCullough, L., Bhatia, M., Ulvenes, P., Berggraf, L., & Osborn, K. (2011). Learning how to rate video-recorded therapy sessions: A practical guide for trainees and advanced clinicians. *Psychotherapy, 48*(2), 127–137.

Office for National Statistics (2018). Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/surveysusingthe4officefornationalstatisticspersonalwellbeingquestions

Psychological Science (2018, February 28). *Measurement matters.* Retrieved from https://www.psychologicalscience.org/observer/measurement-matters

Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, *91*(5), 1266-1271.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural" natural experiments" in economics. *Journal of Economic Literature*, *38*(4), 827-874.

Ross, L. P., & Inagaki, T. K. (2022). Recalling prior experiences with a close other can fulfill the need for social connection. *Emotion*.

Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of personality and social psychology*, *79*(2), 286.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. *Well-being: The foundations of hedonic psychology*, *7*, 61-84.

Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important?. *BMJ: British Medical Journal*, *316*(7126), 201.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual review of psychology*, *66*, 487-518.

Spielberger, C. D. (1983). State-trait anxiety inventory for adults.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, *166*(10), 1092-1097.

Stawarczyk, D., Majerus, S., Van der Linden, M., & D'Argembeau, A. (2012). Using the daydreaming frequency scale to investigate the relationships between mind-wandering, psychological well-being, and present-moment awareness. *Frontiers in psychology*, *3*, 363.

Steimer, T. (2022). The biology of fear-and anxiety-related behaviors. *Dialogues in clinical neuroscience*.

Stone, D. (1999). Learning lessons and transferring policy across time, space and disciplines. *Politics*, *19*(1), 51-59.

Power, M. J., & Tarsia, M. (2007). Basic and complex emotions in depression and anxiety. *Clinical Psychology & Psychotherapy*, *14*(1), 19-31.

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical psychology review*, *31*(7), 1156-1168.

Tomkins, S. (1962). Affect, imagery, consciousness (Vol. 1). New York: Springer

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: a systematic review of the literature. *Psychotherapy and psychosomatics*, *84*(3), 167-176.

van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour*, *5*(5), 631-652.

Westerhof, G. J., & Keyes, C. L. (2010). Mental illness and mental health: The two continua model across the lifespan. *Journal of adult development*, *17*(2), 110-119.

Wood, A. M., & Tarrier, N. (2010). Positive clinical psychology: A new vision and strategy for integrated research and practice. *Clinical psychology review*, *30*(7), 819-829.

Yap, S. C., Wortman, J., Anusic, I., Baker, S. G., Scherer, L. D., Donnellan, M. B., & Lucas, R. E. (2017). The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal of Personality and Social Psychology*, *113*(6), 939.

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, *11*(4), 705.

6. Critical discussion references

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: anticipated aloneness reduces intelligent thought. *Journal of personality and social psychology*, *83*(4), 817.

Beck, A. T., & Haigh, E. A. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual review of clinical psychology*, *10*(1), 1-24.

Bonner, C., Tuckerman, J., Kaufman, J., Costa, D., Durrheim, D. N., Trevena, L., ... & Danchin, M. (2021). Comparing inductive and deductive analysis techniques to understand health service implementation problems: a case study of childhood vaccination barriers. *Implementation Science Communications*, *2*, 1-9.

Cacioppo, J. T., Hughes, M. E., Waite, L. J., Hawkley, L. C., & Thisted, R. A. (2006). Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. *Psychology and aging*, *21*(1), 140.

Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. WW Norton & Company.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, *16*(3), 031001.

Crocker, L. D., Heller, W., Warren, S. L., O'Hare, A. J., Infantolino, Z. P., & Miller, G. A. (2013). Relationships among cognition, emotion, and motivation: implications for intervention and neuroplasticity in psychopathology. *Frontiers in human neuroscience*, *7*, 261.

Fava, G. A., Ruini, C., Rafanelli, C., Finos, L., Salmaso, L., Mangelli, L., & Sirigatti, S. (2005). Well-being therapy of generalized anxiety disorder. *Psychotherapy and psychosomatics*, *74*(1), 26-30.

Ford, B. Q., Dmitrieva, J. O., Heller, D., Chentsova-Dutton, Y., Grossmann, I., Tamir, M., ... & Mauss, I. B. (2015). Culture shapes whether the pursuit of happiness predicts higher or lower well-being. *Journal of Experimental Psychology: General*, *144*(6), 1053.

Gan, D. Z., McGillivray, L., Larsen, M. E., Christensen, H., & Torok, M. (2022). Technology-supported strategies for promoting user engagement with digital mental health interventions: A systematic review. *Digital health*, *8*, 20552076221098268.

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior therapy*, *44*(2), 180-198.

Lee Duckworth, A., Steen, T. A., & Seligman, M. E. (2005). Positive psychology in clinical practice. *Annu. Rev. Clin. Psychol.*, *1*, 629-651.

Layard, R., & Clark, D. M. (2014). *Thrive: The power of evidence-based psychological therapies*. Penguin UK.

McNally, R. J. (2021). Network analysis of psychopathology: controversies and challenges. *Annual review of clinical psychology*, *17*, 31-53.

Michie, S., Carey, R. N., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P., & Connell, L. E. (2018). From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. *Annals of behavioral medicine*, *52*(6), 501-512.

Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., ... & Duckworth, A. L. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, *600*(7889), 478-483.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., ... & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *bmj*, *350*.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., ... & UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The lancet*, *392*(10157), 1553-1598.

Perone, S., & Simmering, V. R. (2017). Applications of dynamic systems theory to cognition and development: New frontiers. *Advances in child development and behavior*, *52*, 43-80.

Poincaré, H. (2017). *The three-body problem and the equations of dynamics: Poincaré's foundational work on dynamical systems theory* (Vol. 443). Springer.

Sun, J., Harris, K., & Vazire, S. (2020). Is well-being associated with the quantity and quality of social interactions?. *Journal of Personality and Social Psychology*, *119*(6), 1478.

Thornicroft, G., & Slade, M. (2014). New trends in assessing the outcomes of mental health interventions. *World Psychiatry*, *13*(2), 118-124.

Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2008). The magnitude of client and therapist working alliance ratings. *Psychotherapy: Theory, Research, Practice, Training*, *45*(4), 546.

Van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour*, *5*(5), 631-652.

Van Bilsen, H. (2018). *Cognitive behaviour therapy in the real world: Back to basics*. Routledge.

Van Valkengoed, A. M., Abrahamse, W., & Steg, L. (2022). To select effective interventions for pro-environmental behaviour change, we need to consider determinants of behaviour. *Nature Human Behaviour*, 1-11.

Walker, E. M., Mwaria, M., Coppola, N., & Chen, C. (2014). Improving the replication success of evidence-based interventions: Why a preimplementation phase matters. *Journal of Adolescent Health*, *54*(3), S24-S28.

Walker, S. (2017). *Learning theory and behaviour modification*. Routledge.

Wight, D., Wimbush, E., Jepson, R., & Doi, L. (2016). Six steps in quality intervention development (6SQuID). *J Epidemiol Community Health*, *70*(5), 520-525.

Whiston, A., Bockting, C. L., & Semkovska, M. (2019). Towards personalising treatment: a systematic review and meta-analysis of face-to-face efficacy moderators of cognitive-behavioral therapy and interpersonal psychotherapy for major depressive disorder. *Psychological Medicine*, *49*(16), 2657-2668.

Yap, S. C., Wortman, J., Anusic, I., Baker, S. G., Scherer, L. D., Donnellan, M. B., & Lucas, R. E. (2017). The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal of Personality and Social Psychology*, *113*(6), 939.

Yanos, P. T., Rosenfield, S., & Horwitz, A. V. (2001). Negative and supportive social interactions and quality of life among persons diagnosed with severe mental illness. *Community Mental Health Journal*, *37*(5), 405-419.

Young, J. E., Rygh, J. L., Weinberger, A. D., & Beck, A. T. (2014). Cognitive therapy for depression.

7. Paper 1, References

Allin, P. Measuring societal wellbeing. *Economic & Labour Market Review*, 1(10), 46-52 (2007).

Diener, E., Lucas, R. E., & Oishi, S. Subjective well-being: The science of happiness and life satisfaction. *Handbook of positive psychology*, ***2*,** 63-73 (2002).

Dolan, P. *Happiness by design: Finding pleasure and purpose in everyday life* (Penguin, (2014).

Diener, E., Sandvik, E., & Pavot, W. Happiness is the frequency, not the intensity, of positive versus negative affect. In *Assessing well-being*. 213-231 (Springer, Dordrecht, 2009).

Dolan, P. *Happiness by design: Finding pleasure and purpose in everyday life* (Penguin, (2014).

Dolan, P., & Metcalfe, R. Measuring subjective wellbeing: Recommendations on measures for use by national governments. *Journal of social policy*, *41*(2), 409-427 (2012).

Dolan, P., Layard, R., & Metcalfe, R. *Measuring subjective well-being for public policy.* Retrieved from: http://eprints.lse.ac.uk/35420/1/measuring-subjective-wellbeing-for-public-policy.pdf (2011).

Dolan, P., Peasgood, T., & White, M. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economics*, *29*(1), 94-122 (2008).

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191 (2007).

Gloor, P, A., Colladon, A. F., Grippa, F., Budner, P., & Eirich, J. Aristotle said "happiness is a state of activity"—Predicting mood through body sensing with Smartwatches. *Journal of Systems Science and Systems Engineering*, *27*(5), 586-612 (2018).

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776-1780 (2004).

Kristoffersen, I. The metrics of subjective wellbeing: Cardinality, neutrality and additivity. *Economic Record*, *86*(272), 98-123 (2010).

Krueger, A. B., & Stone, A. A. Progress in measuring subjective well-being. *Science*, *346*(6205), 42-43 (2014).

Picard, R. W., & Daily, S. B. (2005, April). Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches* (Vol. 10, No. 1056808.1057115, pp. 2119-2122). New York, NY: ACM.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98-125 (2017).

Røysamb, E., Nes, R. B., Czajkowski, N. O., & Vassend, O. Genetics, personality and wellbeing. A twin study of traits, facets and life satisfaction. *Scientific Reports*, *8*(1), 1-13 (2018).

Seresinhe, C. I., Preis, T., MacKerron, G., & Moat, H. S. Happiness is greater in more scenic locations. *Scientific reports*, *9*(1), 1-11 (2019).

Stone, A, A., Shiffman, S, S., DeVries, M, W. "Ecological Momentary Assessment", in *Well-being: The Foundations of Hedonic Psychology,* D. Kahneman, E. Diener and N. Schwarz, Eds. 26-39 (Russell-Sage, 1999)

Stone, A. A., & Shiffman, S. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*(3), 236-243 (2002).

Slavish, D. C., Sliwinski, M. J., Smyth, J. M., Almeida, D. M., Lipton, R. B., Katz, M. J., & Graham-Engeland, J. E. Neuroticism, rumination, negative affect, and sleep: Examining between-and within-person associations. *Personality and Individual Differences*, *123*, 217-222 (2018).

Tao, J., & Tan, T. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*. 981-995. Springer, Berlin, Heidelberg (2005, October)

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. Predicting the duration of emotional experience: Two experience sampling studies. *Emotion*, *9*(1), 83 (2009).

Voukelatou, V., Gabrielli, L., Miliou, I., Cresci, S., Sharma, R., Tesconi, M., & Pappalardo, L. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics*, 1-31 (2020).

8. Paper 2, References

Agresti, A. (2018). An introduction to categorical data analysis (Vol. 2.). Wiley

Ando, M. (2002). Relationships among mental health, coping styles, and mood. *Psychological reports*, *90*(2), 606-612.

Andrews, P., & Meyer, R. G. (2003). Marlowe–Crowne social desirability scale and short form C: forensic norms. *Journal of clinical psychology*, *59*(4), 483-492.

Baddeley, A., Banse, R., Huang, Y. M., & Page, M. (2012). Working memory and emotion: Detecting the hedonic detector. *Journal of Cognitive Psychology*, *24*(1), 6-16.

Baker, R. C., & Guttfreund, D. O. (1993). The effects of written autobiographical recollection induction procedures on mood. *Journal of Clinical Psychology*, *49*(4), 563-568.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1-23.

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science*, *8*(1), 10-14.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review*, *8*(1), 77-100.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry*, *4*(6), 561-571.

Becker, M. W., & Leinenger, M. (2011). Attentional selection is biased toward mood-congruent stimuli. *Emotion*, *11*(5), 1248.

Belzer, K., & Schneier, F. R. (2004). Comorbidity of anxiety and depressive disorders: issues in conceptualization, assessment, and treatment. *Journal of Psychiatric Practice®*, *10*(5), 296-306.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, *57*(1), 289-300.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*, *11*(2), e0148037.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Brahler, E., Schumacher, J., & Strauss, B. (2002). *Diagnostische verfahren in der psychotheraphie.* Hogrefer, Verlag fur Psychologie.

Bowen, R. C., Wang, Y., Balbuena, L., Houmphan, A., & Baetz, M. (2013). The relationship between mood instability and depression: implications for studying and treating depression. *Medical Hypotheses*, *81*(3), 459-462.

Bower, G. H. (1981). Mood and memory. *American psychologist*, *36*(2), 129.

Chalder, M., Wiles, N. J., Campbell, J., Hollinghurst, S. P., Haase, A. M., Taylor, A. H., ... & Lewis, G. (2012). Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *Bmj*, *344*.

Chentsova-Dutton, Y. E., Choi, E., Ryder, A. G., & Reyes, J. (2015). "I felt sad and did not enjoy life": Cultural context and the associations between anhedonia, depressed mood, and momentary emotions. *Transcultural psychiatry*, *52*(5), 616-635.

Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behaviour Research and Therapy*, *5*(1), 27-49.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic impli- cations. Journal of Abnormal Psychology, 100, 316–336.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Dobson, K. S. (1985). The relationship between anxiety and depression. *Clinical Psychology Review*, *5*(4), 307-324.

Ekman, P., 1999. Basic Emotions. In: Dalgleish, T., Power, M., (eds.). *Handbook of Cognition and Emotion.* Sussex, UK: John Wiley & Sons

Ferrer, R. A., Grenen, E. G., & Taber, J. M. (2015). Effectiveness of internet-based affect induction procedures: A systematic review and meta-analysis. *Emotion*, *15*(6), 752.

Firth, J., Siddiqi, N., Koyanagi, A., Siskind, D., Rosenbaum, S., & Galletly, C. (2019). A blueprint for protecting physical health in people with mental illness: directions for health promotion, clinical services and future research. *Lancet Psychiatry*.

Forgas, J. P. (2013). Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. *Current Directions in Psychological Science*, *22*(3), 225-232.

Forrest, S. J., Siegert, R. J., Krägeloh, C. U., Landon, J., & Medvedev, O. N. (2021). Generalizability theory distinguishes between state and trait anxiety. *Psychological Assessment*.

Firth, J., Siddiqi, N., Koyanagi, A. I., Siskind, D., Rosenbaum, S., Galletly, C., ... & Stubbs, B. (2019). The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *The Lancet Psychiatry*, *6*(8), 675-712.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions.

Gaudry, E., Vagg, P., & Spielberger, C. D. (1975). Validation of the state-trait distinction in anxiety research. *Multivariate behavioral research*, *10*(3), 331-341.

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality Reconsidered: Diversity in Making Meaning of Facial Expressions. *Current Directions in Psychological Science, 27*(4), 211-219.

Geschwind, N., Peeters, F., Drukker, M., van Os, J., & Wichers, M. (2011). Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: a randomized controlled trial. *Journal of consulting and clinical psychology*, *79*(5), 618.

Göritz, A. S., & Moser, K. (2006). Web-based mood induction. *Cognition and Emotion*, *20*(6), 887-896.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.

Izard, Carroll E. (1977), Human Emotions, New York: Plenum Press, p.94.

Jallais, C., & Gilet, A. L. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior research methods*, *42*(1), 318-325.

Kelly, C., & Sharot, T. (2023). Knowledge-Seeking Reflects and Shapes Well-Being.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of consulting and clinical psychology*, *73*(3), 539.

Krauth-Gruber, S., & Ric, F. (2000). Affect and stereotypic thinking: A test of the mood-and-general-knowledge model. *Personality & Social Psychology Bulletin*, 26, 1587-1597. doi:10.1177/ 01461672002612012

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of affective disorders*, *114*(1-3), 163-173.

Larsen, R. J., & Ketelaar, T. (1989). Extraversion, neuroticism and susceptibility to positive and negative mood induction procedures. *Personality and individual Differences*, *10*(12), 1221-1228.

Laux, L., Hock, M., Bergner-Kother, R., Hodapp, V., & Renner, K. H. (2013). Das State-Trait-Angst-Depressions-Inventar (STADI). Gottingen: Hogrefe.

Lench, H., Flores, S., & Bench, S. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, *137*, 834–855. doi:10.1037/a0024244

Lubin, B. (1965). Adjective checklists for measurement of depression. *Archives of general psychiatry*, *12*(1), 57-62.

McDaid, D., Park, A. L., & Wahlbeck, K. (2019). The economic case for the prevention of mental illness. *Annual Review of Public Health*, *40*, 373-389.

Office for National Statistics (2018). Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/surveysusingthe4officefornationalstatisticspersonalwellbeingquestions

Open Science Framework, January 2021, Retrieved from (link not given as per masked review guidelines but can provide at later date)

Patel, R., Lloyd, T., Jackson, R., Ball, M., Shetty, H., Broadbent, M., ... & Taylor, M. (2015). Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ open*, *5*(5), e007504.

Payton, A. R. (2009). Mental health, mental illness, and psychological distress: same continuum or distinct phenomena?. *Journal of health and Social Behavior*, *50*(2), 213-227.

Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion*, *6*(3), 383.

Prkachin, K. M., Williams-Avery, R. M., Zwaal, C., & Mills, D. E. (1999). Cardiovascular changes during induced emotion: An application of Lang's theory of emotional imagery. *Journal of psychosomatic research*, *47*(3), 255-267.

Robinson, O. J., & Sahakian, B. J. (2009). A double dissociation in the roles of serotonin and mood in healthy subjects. *Biological psychiatry*, *65*(1), 89-92.

Robinson, O. J., Grillon, C., & Sahakian, B. J. (2012). The mood induction task: A standardized, computerized laboratory procedure for altering mood state in humans.

Rönkkö, M., & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 1094428120968614.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of clinical psychology*, *62*(1), 123-146.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, *19*(1), 80-110.

Shin, K. E., Newman, M. G., & Jacobson, N. C. (2021). Emotion network density is a potential clinical marker for anxiety and depression: Comparison of ecological momentary assessment and daily diary. *British Journal of Clinical Psychology*.

Siedlecka, E., & Denson, T. F. (2019). Experimental methods for inducing basic emotions: A qualitative review. *Emotion Review*, *11*(1), 87-97.

Siedlecka, E., Capper, M. M., & Denson, T. F. (2015). Negative emotional events that people ruminate about feel closer in time. *PloS one*, *10*(2), e0117105.

Spielberger, C. D. (1980). Test manual for the State-Trait Anxiety Inventory-STAI form Y. *Y Consulting Psychologist Press, Palo Alto, California*.

Spielberger, C. D. (1989). *State-Trait Anxiety Inventory: Bibliography* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Targum, S. D., Sauder, C., Evans, M., Saber, J. N., & Harvey, P. D. (2021). Ecological momentary assessment as a measurement tool in depression trials. *Journal of psychiatric research*, *136*, 256-264.

Van Agteren, J., Iasiello, M., Lo, L., Bartholomaeus, J., Kopsaftis, Z., Carey, M., & Kyrios, M. (2021). A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. *Nature Human Behaviour*, *5*(5), 631-652.

Van der Does, W. (2002). Different types of experimentally induced sad mood? *Behavior Therapy*, *33*(4), 551-561.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of abnormal psychology*, *104*(1), 3.

Watson, J. C., Goldman, R. N., & Greenberg, L. S. (2011). Contrasting Two Clients in Emotion-Focused Therapy for Depression 1: The Case of" Tom,"" Trapped in the Tunnel". *Pragmatic Case Studies in Psychotherapy*, *7*(2), 268-304.

Yaden, D. B., & Haybron, D. M. (2022). The emotional state assessment tool: a brief, philosophically informed, and cross-culturally sensitive measure. *The Journal of Positive Psychology*, 1-15.

9. Paper 3, References

Alexopoulos, T., & Ric, F. (2007). The evaluation-behavior link: Direct and beyond valence. *Journal of Experimental Social Psychology*, *43*, 1010-1016.

Amini, S., Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2012). Fixed vs random: The Hausman test four decades later. In *Essays in honor of Jerry Hausman*. Emerald Group Publishing Limited.

Asnaani, A., Rinck, M., Becker, E., & Hofmann, S. G. (2014). The effects of approach–avoidance modification on social anxiety disorder: A pilot study. *Cognitive Therapy and Research*, *38*(2), 226-238.

Bacon, A. M., Krupić, D., Caki, N., & Corr, P. J. (2022). Emotional and Behavioral Responses to COVID-19. *European Psychologist*.

Baumeister, R. F., Twenge, J. M., & Nuss, C. K. (2002). Effects of social exclusion on cognitive processes: anticipated aloneness reduces intelligent thought. *Journal of personality and social psychology*, *83*(4), 817.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: series B (Methodological)*, *57*(1), 289-300.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS One*, *11*(2), e0148037.

BMJ, *Covid-19: Experts recommend two week circuit break to reduce transmission in the UK* (2020, October 16). Retrieved February 30, 2022, from https://doi.org/10.1136/bmj.m4038

Cacioppo, J. T., Hawkley, L. C., Ernst, J. M., Burleson, M., Berntson, G. G., Nouriani, B., & Spiegel, D. (2006). Loneliness within a nomological net: An evolutionary perspective. *Journal of research in personality*, *40*(6), 1054-1085.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and emotion*, *30*(2), 105-110.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology*, *67*(2), 319.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*, 215-224.

Croissant, Yves & Millo, Giovanni. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*. 27. 1-43. 10.18637/jss.v027.i02.

De Lange*, A. H., Taris, T. W., Kompier, M. A., Houtman, I. L., & Bongers, P. M. (2004). The relationships between work characteristics and mental health: Examining normal, reversed and reciprocal relationships in a 4-wave study. *Work & Stress*, *18*(2), 149-166.

Dolan, P., Layard, R., O'Donnell, G., Delaney, L., Krekel, C., Sanders, J., ... & Kudrna, L. (2021). How a focus on wellbeing can help us make better policy decisions. *LSE COVID-19 Blog*.

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L., Recchia, G., Van Der Bles, A. M., ... & Van Der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, *23*(7-8), 994-1006.

Econometrics with R (pp. 135-152). Springer. Retrieved from https://www.econometrics-with-r.org/10-rwpd.html

Elliot, A. J., & Friedman, R. (2017). Approach—Avoidance: A Central Characteristic 01 Personal Goals. In *Personal project pursuit* (pp. 3-50). Psychology Press.

Fancourt, D., Steptoe, A., & Bu, F. (2021). Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. *The Lancet Psychiatry*, *8*(2), 141-149.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231-1236.

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests–An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*, 95-102.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156-168.

Ganesan, B., Al-Jumaily, A., Fong, K. N., Prasad, P., Meena, S. K., & Tong, R. K. Y. (2021). Impact of coronavirus disease 2019 (COVID-19) outbreak quarantine, isolation, and lockdown policies on mental health and suicide. Frontiers in psychiatry, 471.

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*, *67*(8), 850-857.

Greater London Authority. (2021). Covid-19 restrictions timeseries. London Datastore. Retrieved from https://data.london.gov.uk/dataset/covid-19-restrictions-timeseries

Grieve, R., Indian, M., Witteveen, K., Tolan, G. A., & Marrington, J. (2013). Face-to-face or Facebook: Can social connectedness be derived online?. *Computers in Human Behavior*, *29*(3), 604-609.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological inquiry*, *26*(1), 1-26.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the Liebowitz social anxiety scale. *Psychological Medicine*, *29*(1), 199-212.

Heuer, K., Rinck, M., & Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: The approach–avoidance task. *Behaviour Research and Therapy*, *45*(12), 2990-3001.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., ... & Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*(7820), 262-267.

Huntsinger, J. R. (2013). Incidental experiences of affective coherence and incoherence influence persuasion. *Personality and Social Psychology Bulletin*, *39*(6), 792-802.

Kelloway, E. K., & Francis, L. (2013). Longitudinal research and data analysis.

Krieglmeyer, R., De Houwer, J., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review*, *5*(3), 280-284.

Krishna, A., Rodrigues, J., Mitschke, V., & Eder, A. B. (2021). Self-reported mask-related worrying reduces relative avoidance bias toward unmasked faces in individuals with low Covid19 anxiety syndrome. *Cognitive Research: Principles and Implications, 6*(1), 1-9.

Krupić, D., Žuro, B., & Krupić, D. (2021). Big Five traits, approach-avoidance motivation, concerns and adherence with COVID-19 prevention guidelines during the peak of pandemic in Croatia. *Personality and Individual Differences, 179*, 110913.

Krypotos, A. M., Arnaudova, I., Effting, M., Kindt, M., & Beckers, T. (2015). Effects of approach-avoidance training on the extinction and return of fear responses. *PloS One*, *10*(7), e0131581.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International Affective Picture System (IAPS): Instruction manual and affective ratings. The Center for Research in Psychophysiology, University of Florida.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*, 1377–1388.

Loades, M. E., Chatburn, E., Higson-Sweeney, N., Reynolds, S., Shafran, R., Brigden, A., ... & Crawley, E. (2020). Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(11), 1218-1239.

Loijen, A., Vrijsen, J. N., Egger, J. I., Becker, E. S., & Rinck, M. (2020). Biased approach-avoidance tendencies in psychopathology: A systematic review of their assessment and modification. *Clinical Psychology Review*, *77*, 101825.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high quality, realistic picture database. *Behavior Research Methods, 46,* 596-610. doi: 10.3758/s13428-013-0379-1

Marroquín, B., Vine, V., & Morgan, R. (2020). Mental health during the COVID-19 pandemic: Effects of stay-at-home policies, social distancing behavior, and social resources. *Psychiatry Research*, *293*, 113419.

Matias, T., Dominski, F. H., & Marks, D. F. (2020). Human needs in COVID-19 isolation. *Journal of Health Psychology*, *25*(7), 871-882.

McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion, 20* (1), 1–9.

Monni, A., Olivier, E., Morin, A. J. S., Belardinelli, M. O., Mulvihill, K., & Scalas, L. F. (2020). Approach and avoidance in Gray's, Higgins', and Elliot's perspectives: A theoretical comparison and integration of approach-avoidance in motivated behavior. *Personality and Individual Differences*, *166*, 110163.

Neumann, R., Förster, J., & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation* (pp. 371-391). London, UK: Lawrence Erlbaum Associates.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic bulletin & review*, *26*(5), 1596-1618.

Pancani, L., Marinucci, M., Aureli, N., & Riva, P. (2021). Forced social isolation and mental health: a study on 1,006 Italians under COVID-19 lockdown. Frontiers in Psychology, 12, 663799.

Jin, L. (2017) Panel data analysis: Lecture note. Retrieved from https://www.fsb.miamioh.edu/lij14/411_note_panel.pdf

Peschard, V., & Philippot, P. (2017). Overestimation of threat from neutral faces and voices in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *57*, 206-211

PLM package, *Cran r-project* (ND), Retrieved November 20, 2022, from, https://cran.r-project.org/web/packages/plm/vignettes/A_plmPackage.html#nlme

Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, *91*(5), 1266-1271.

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*(4), 827-874.

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 105-120.

Schmitz, M., & Wentura, D. (2012). Evaluative priming of naming and semantic categorization responses revisited: A mutual facilitation explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 984–1000.

Sharbanee, J. M., Hu, L., Stritzke, W. G., Wiers, R. W., Rinck, M., & MacLeod, C. (2014). The effect of approach/avoidance training on alcohol consumption is mediated by change in alcohol action tendency. *PloS One*, *9*(1), e85855.

Sharot, T. (2012). *The Optimism Bias: Why we're wired to look on the bright side*. Hachette UK.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*(1), 487-510.

Seibt, B., Neumann, R., Nussinson, R., & Strack, F. (2008). Movement direction or change in distance? Self- and object-related approach-avoidance motions. *Journal of Experimental Social Psychology, 44*, 713-720.

Sen, K., Prybutok, G., & Prybutok, V. (2022). The use of digital technology for social wellbeing reduces social isolation in older adults: A systematic review. *SSM-population health*, *17*, 101020.

Spielberger, C. D. (1983). State-trait anxiety inventory for adults.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220-247.

Shamblaw, A. L., Rumas, R. L., & Best, M. W. (2021). Coping during the COVID-19 pandemic: Relations with mental health and quality of life. *Canadian Psychology / Psychologie canadienne, 62*(1), 92–100

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy*, 50, 529-536.

Thompson, R. A. (2011). Emotion and emotion regulation: Two sides of the developing coin. *Emotion Review*, *3*(1), 53-61.

University of Bristol. (n.d.) Multilevel models: What and why? Centre for Multilevel Modelling. Retrieved from https://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html

Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *Journal of Population Economics*, *33*(4), 1127-1172.

Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfelder, S. (2010). EmoPics: Subjektive und psychophysiologische Evaluation neuen Bildmaterials für die klinisch-biopsychologische Forschung. *Zeitschrift für Klinische Psychologie und Psychotherapie*, Supplementum 1/10, 77.

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Re-training automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22,* 490-497*.* DOI: 10.1177/0956797611400615

World Health Organisation, *Coronavirus Dashboard* (ND). Retrieved January 20, 2022, from, https://covid19.who.int

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q. P., & Lillard Jr, J. W. (2014). A study of effects of multicollinearity in the multivariable analysis. *International journal of applied science and technology*, *4*, 9-19.

10. Paper 4, References

Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PloS one*, *16*(3), e0248152.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, *15*(6), 713-724.

Beames, J. R., Kikas, K., & Werner-Seidler, A. (2021). Prevention and early intervention of depression in young people: an integrated narrative review of affective awareness and Ecological Momentary Assessment. *BMC psychology*, *9*(1), 113.

Catuara-Solarz, S., Skorulski, B., Estella-Aguerri, I., Avella-Garcia, C. B., Shepherd, S., Stott, E., ... & Dix, S. (2022). The Efficacy of "Foundations," a Digital Mental Health App to Improve Mental Well-being During COVID-19: Proof-of-Principle Randomized Controlled Trial. *JMIR mHealth and uHealth*, *10*(7), e30976.

Cloninger, C. R. (2006). The science of well-being: an integrated approach to mental health and its disorders. *World psychiatry*, *5*(2), 71.

Coffey, K. A., Hartman, M., & Fredrickson, B. L. (2010). Deconstructing mindfulness and constructing mental health: Understanding mindfulness and its mechanisms of action. *Mindfulness*, *1*, 235-253.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385-396.

Collins, R. L., Morsheimer, E. T., Shiffman, S., Paty, J. A., Gnys, M., & Papandonatos, G. D. (1998). Ecological momentary assessment in a behavioral drinking moderation training program. *Experimental and Clinical Psychopharmacology*, *6*(3), 306.

Conner, T. S., & Reid, K. A. (2012). Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science*, *3*(3), 315-323.

Cosci, F., & Fava, G. A. (2021). When anxiety and depression coexist: the role of differential diagnosis using clinimetric criteria. *Psychotherapy and Psychosomatics*, *90*(5), 308-317.

Demiralp, E., Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Barrett, L. F., ... & Jonides, J. (2012). Feeling blue or turquoise? Emotional differentiation in major depressive disorder. *Psychological science*, *23*(11), 1410-1416.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of personality and social psychology*, *47*(5), 1105.

Dolan, P., Kudrna, L., & Testoni, S. (2017). Definition and measures of subjective wellbeing. *Centre for Economic Performance (Discussion paper 3)*, 1-9.

Edwards, E. R., & Wupperman, P. (2017). Emotion regulation mediates effects of alexithymia and emotion differentiation on impulsive aggressive behavior. *Deviant behavior*, *38*(10), 1160-1171.

Enkema, M. C., McClain, L., Bird, E. R., Halvorson, M. A., & Larimer, M. E. (2020). Associations between mindfulness and mental health outcomes: A systematic review of ecological momentary assessment research. *Mindfulness*, *11*, 2455-2469.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287-298.

Fogarty, F. A., Lu, L. M., Sollers, J. J., Krivoschekov, S. G., Booth, R. J., & Consedine, N. S. (2015). Why it pays to be mindful: Trait mindfulness predicts physiological recovery from emotional stress and greater differentiation among negative emotions. *Mindfulness*, *6*, 175-185.

Frewen, P. A., Evans, E. M., Maraj, N., Dozois, D. J., & Partridge, K. (2008). Letting go: Mindfulness and negative automatic thinking. *Cognitive therapy and research*, *32*(6), 758-774.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156-168.

Goldstein, M. D., & Strube, M. J. (1994). Independence revisited: The relation between positive and negative affect in a naturalistic setting. *Personality and Social Psychology Bulletin*, *20*(1), 57-64.

Heisenberg, W. (1930). *The physical principles of the quantum theory*. University of Chicago Press, Chicago.

Helliwell, J. F., Huang, H., Wang, S., & Norton, M. (2021). World happiness, trust and deaths under COVID-19. *World happiness report*, *2021*, 13-57.

Heron, K. E., & Smyth, J. M. (2013). Is intensive measurement of body image reactive? A two-study evaluation using Ecological Momentary Assessment suggests not. *Body Image*, *10*(1), 35-44.

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of consulting and clinical psychology*, *78*(2), 169.

Holmes, E. A., Craske, M. G., & Graybiel, A. M. (2014). Psychological treatments: A call for mental-health science. *Nature*, *511*(7509), 287-289.

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. *Psychology of addictive behaviors*, *16*(3), 205.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, *306*(5702), 1776-1780.

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, *24*(1), 10-16.

Kashdan, T. B., Ferssizidis, P., Collins, R. L., & Muraven, M. (2010). Emotion differentiation as resilience against excessive alcohol use: An ecological momentary assessment in underage social drinkers. *Psychological science*, *21*(9), 1341-1347.

Kauer, S. D., Reid, S. C., Crooke, A. H. D., Khor, A., Hearps, S. J. C., Jorm, A. F., ... & Patton, G. (2012). Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *Journal of medical Internet research*, *14*(3), e1858.

Kramer, I., Simons, C. J., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., ... & Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry*, *13*(1), 68-77.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932-932.

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological science*, *23*(10), 1086-1091.

Koydemir, S., Sökmez, A. B., & Schütz, A. (2021). A meta-analysis of the effectiveness of randomized controlled positive psychological interventions on subjective and psychological well-being. *Applied Research in Quality of Life*, *16*, 1145-1185.

Kreitler, S. (2018). The meaning profiles of anxiety and depression: Similarities and differences in two age groups. *Cognition and Emotion*, *32*(7), 1499-1513.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509-515.

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *Journal of clinical psychology*, *67*(1), 99-110.

Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychology*, *17*(1), 48.

Louet, S. (2015). Mental health: the mindful way. *Nature*, *527*(7579), 553-554.

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2021). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of personality and social psychology*, *120*(3), 816.

Magnan, R. E., Köblitz, A. R., McCaul, K. D., & Dillard, A. J. (2013). Self-monitoring effects of ecological momentary assessment on smokers' perceived risk and worry. *Psychological assessment*, *25*(2), 416.

McCarthy, D. E., Minami, H., Yeh, V. M., & Bold, K. W. (2015). An experimental investigation of reactivity to ecological momentary assessment frequency among adults trying to quit smoking. *Addiction*, *110*(10), 1549-1560.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry*, *17*(2), 123-132.

Myin-Germeys, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. *Current opinion in psychiatry*, *29*(4), 258-263.

OECD European Union, 2018. Accessed: https://www.oecd-ilibrary.org/docserver/health_glance_eur-2018-4-en.pdf?expires=1678473918&id=id&accname=ocid71015720&checksum=E3913E5ECE6A97C3FC36330E62F0CF58

Office for National Statistics. (2019). Surveys using the 4 Office for National Statistics personal well-being questions: Methodology. Retrieved March 25, 2023, from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/surveysusingthe4officefornationalstatisticspersonalwellbeingquestions

Open Science Foundation, 2019: https://osf.io/yt745

Open Science Foundation, 2021: https://osf.io/hvtf8

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., ... & UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The lancet*, *392*(10157), 1553-1598.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological science*, *8*(3), 162-166.

Pincus, A. L., & Ansell, E. B. (2013). Interpersonal theory of personality.

Pomerantz, A. M., & Rose, P. (2014). Is depression the past tense of anxiety? An empirical study of the temporal distinction. *International Journal of Psychology*, *49*(6), 446-452.

Price, C. J., & Hooven, C. (2018). Interoceptive awareness skills for emotion regulation: Theory and approach of mindful awareness in body-oriented therapy (MABT). *Frontiers in psychology*, *9*, 798.

Schueller, S. M., Aguilera, A., & Mohr, D. C. (2017). Ecological momentary interventions for depression and anxiety. *Depression and anxiety*, *34*(6), 540-545.

Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of clinical psychology*, *65*(5), 467-487.

Shipp, A. J., & Aeon, B. (2019). Temporal focus: Thinking about the past, present, and future. *Current Opinion in Psychology*, *26*, 37-43.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual review of psychology*, *66*, 487-518.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, *166*(10), 1092-1097.

Stawarczyk, D., Majerus, S., Van der Linden, M., & D'Argembeau, A. (2012). Using the daydreaming frequency scale to investigate the relationships between mind-wandering, psychological well-being, and present-moment awareness. *Frontiers in psychology*, *3*, 363.

Stein, K. F., & Corte, C. M. (2003). Ecologic momentary assessment of eating-disordered behaviors. *International Journal of Eating Disorders*, *34*(3), 349-360.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. *Annals of behavioral medicine*.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*(3), 236-243.

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: a systematic review of the literature. *Psychotherapy and psychosomatics*, *84*(3), 167-176.

Weiss, L. A., Westerhof, G. J., & Bohlmeijer, E. T. (2016). Can we increase psychological well-being? The effects of interventions on psychological well-being: A meta-analysis of randomized controlled trials. *PloS one*, *11*(6), e0158092.

Wichers, M., Simons, C. J. P., Kramer, I. M. A., Hartmann, J. A., Lothmann, C., Myin-Germeys, I., & Van Os, J. (2011). Momentary assessment technology as a tool to help patients with depression help themselves. *Acta psychiatrica scandinavica*, *124*(4), 262-272.

Widdershoven, R. L., Wichers, M., Kuppens, P., Hartmann, J. A., Menne-Lothmann, C., Simons, C. J., & Bastiaansen, J. A. (2019). Effect of self-monitoring through experience sampling on emotion differentiation in depression. *Journal of affective disorders*, *244*, 71-77.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current directions in psychological science*, *14*(3), 131-134.

White, M. P., & Dolan, P. (2009). Accounting for the richness of daily activities. *Psychological science*, *20*(8), 1000-1008.

11. Paper 1, Appendix

This PDF file includes:

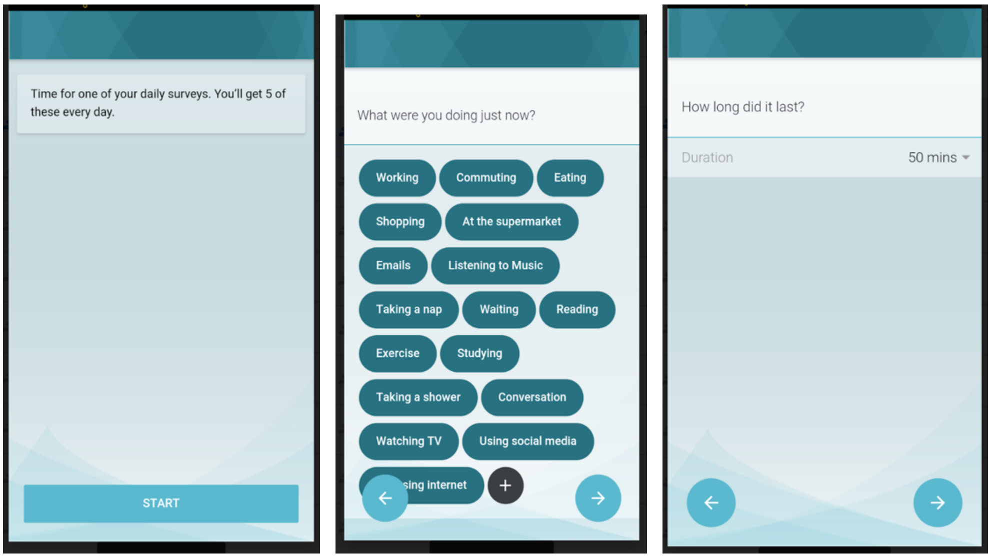
Materials (more detail)

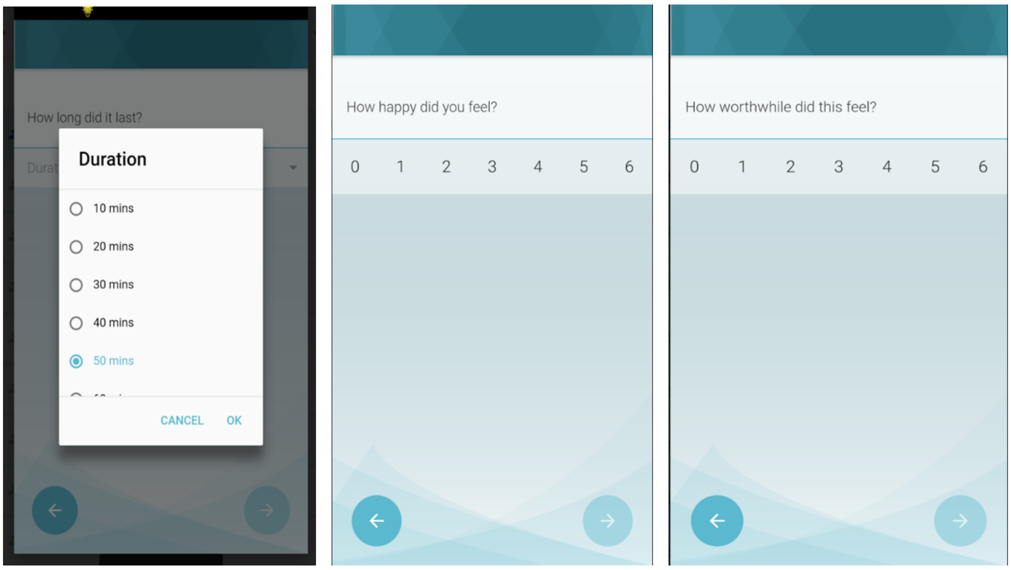
Statistical Analysis (more detail)

Tables S1-S14

Materials

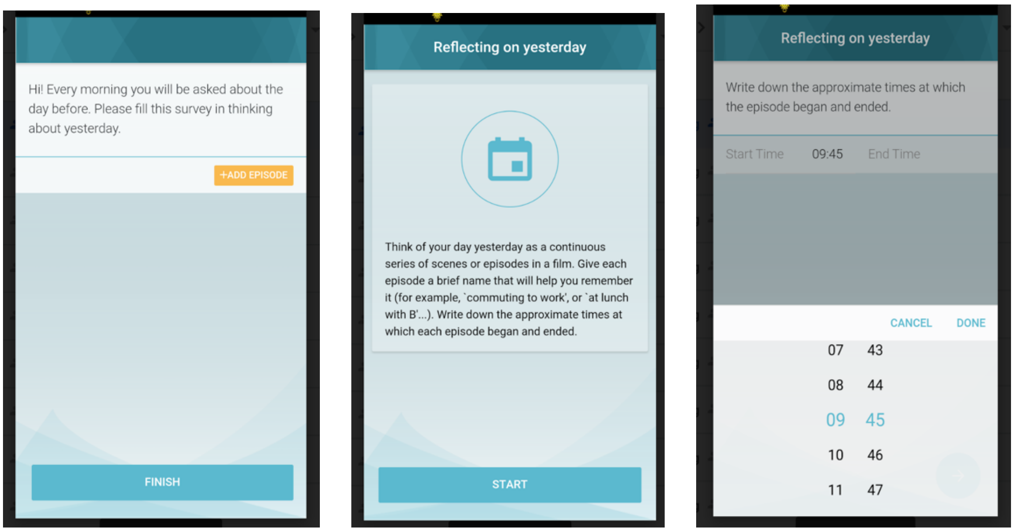
Ecological Momentary Assessment

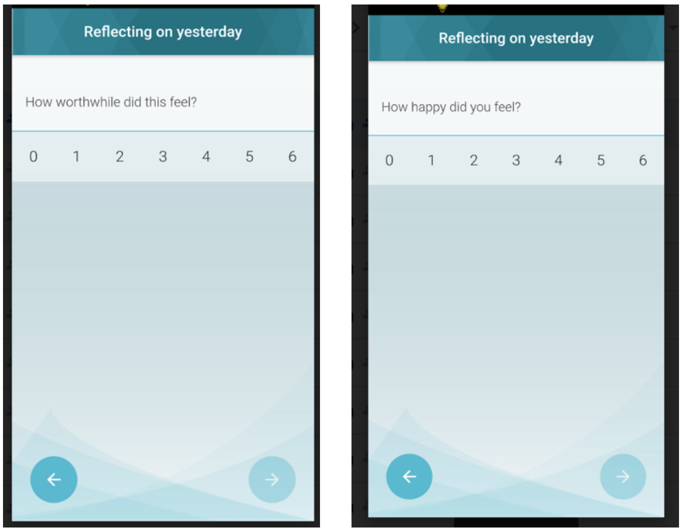




Screenshots of the Ecological Momentary Assessment measure taken from the Reflections app that was used to collect our data

Day Reconstruction Method





Screenshots of the Day Reconstruction Method measure taken from the Reflections app that was used to collect our data

Statistical Analysis

Robustness check for t-tests

For robustness, we used four different formulas for calculating SWB: 1) total average SWB scores aggregated over the full length of the 2-3 week studies (Total SWB), 2) total average SWB scores aggregated over the full length of the 2-3 week studies weighted by duration (Total SWB weighted), 3) average of daily SWB scores (Daily SWB), 4) average of daily SWB scores weighted by duration (Daily SWB weighted). See below for details.

- **Total SWB** = $\frac{SWB_1 + SWB_2 + \dots + SWB_Q}{Q}\frac{SWB_1 + SWB_2 + \dots + SWB_Q}{Q}$, where SWB_i is the reported wellbeing associated to the i-th activity and Q is the number of questionnaires answered;
- **Total *weighted* SWB** = $\frac{Dur_1\, SWB_1 + Dur_2\, SWB_2 + \dots + Dur_Q\, SWB_Q}{Dur_1 + Dur_2 + \dots + Dur_Q}\frac{Dur_1\, SWB_1 + Dur_2\, SWB_2 + \dots + Dur_Q\, SWB_Q}{Dur_1 + Dur_2 + \dots + Dur_Q}$, where Dur_i is the reported duration of the i-th activity and SWB_i and Q are as above;

- **Daily SWB** =
$$\frac{\frac{(SWB_{1,1}+ \dots + SWB_{Q_1,1})}{Q_1}+\dots+\frac{(SWB_{1,D}+ \dots + SWB_{Q_D,D})}{Q_1}}{D}\frac{\frac{(SWB_{1,1}+ \dots + SWB_{Q_1,1})}{Q_1}+\dots+\frac{(SWB_{1,D}+ \dots + SWB_{Q_D,D})}{Q_1}}{D},$$
where $SWB_{j,i}$ is the reported wellbeing associated to the i-th activity on the j-th day, Qj is the number of questionnaires answered on the j-th day and D is the number of days the study lasted;
- **Daily *weighted* SWB** =
$$\frac{\frac{Dur_{1,1}SWB_{1,1}+\dots+Dur_{Q_1,1}SWB_{Q_1,1}}{Dur_{1,1}+\dots+Dur_{Q_1,1}}+\dots+\frac{Dur_{1,D}SWB_{1,D}+\dots+Dur_{Q_D,D}SWB_{Q_D,D}}{Dur_{1,D}+\dots+Dur_{Q_D,D}}}{D}$$
$$\frac{\frac{Dur_{1,1}SWB_{1,1}+\dots+Dur_{Q_1,1}SWB_{Q_1,1}}{Dur_{1,1}+\dots+Dur_{Q_1,1}}+\dots+\frac{Dur_{1,D}SWB_{1,D}+\dots+Dur_{Q_D,D}SWB_{Q_D,D}}{Dur_{1,D}+\dots+Dur_{Q_D,D}}}{D},$$
where $Dur_{i,j}$ is the reported duration of the i-th activity on the j-th day and $SWB_{i,j}$, Qj and D are as in the previous point.

Tables

Table S.1. High variance users - duration

Mean of pairwise differences between average SWB with duration weights and average SWB without duration weights for each type of formula, SWB measure, questionnaire type, and sample. In computing Table S.1, we considered only those users who were in the top quartile of variance of duration of reports. Significant results mean that mean difference is less than a threshold of a tenth of a point. (*** if p-val < 0.001, ** if p-val < 0.01, * if p-val < 0.05).

Mean difference	**Student sample**				**Mixed sample**			
	EMA		**DRM**		**EMA**		**DRM**	
	Total	**Daily**	**Total**	**Daily**	**Total**	**Daily**	**Total**	**Daily**
Happiness	0.059	0.039^{***}	-0.023^{*}	0.003^{***}	0.043	0.011^{***}	-0.077	-0.024^{***}
Worthwhileness	0.107	0.053^{***}	-0.008^{**}	0.011^{***}	0.077	0.039^{***}	-0.035	-0.029^{***}

Table S.2. High variance users - intensity

Mean of pairwise differences between average SWB with duration weights and average SWB without duration weights for each type of formula, SWB measure, questionnaire type, and sample. For Table S.2, we considered those who were in the top quartile of variance of SWB. Significant results mean that mean difference is less than a threshold of a tenth of a point. (*** if p-val < 0.001, ** if p-val < 0.01, * if p-val < 0.05).

Mean difference	**Student sample**				**Mixed sample**			
	EMA		**DRM**		**EMA**		**DRM**	
	Total	**Daily**	**Total**	**Daily**	**Total**	**Daily**	**Total**	**Daily**
Happiness	0.083	0.064	0.042	0.016***	-0.061	-0.011***	-0.073	-0.038*
Worthwhileness	0.223	0.115	0.029	0.035**	0.044	0.013**	0.002	-0.020*

Table S.3. Simulations. Mean of pairwise differences between average happiness with duration weights and average happiness without duration weights.

For each simulation, we randomly selected half of each users' duration and intensity reports. In the average simulation, a single random sample is taken per user. In the low and high correlation simulations, 100 samples were taken per user and the ones with the lowest and highest correlation between happiness and duration were selected. Significant results mean that mean difference is less than a threshold of a tenth of a point. (*** if p-val < 0.001, ** if p-val < 0.01)

Mean difference	Student sample		Mixed sample	
	EMA	DRM	EMA	DRM
Raw data	0.035***	-0.007***	-0.008***	-0.021***
Average simulation	0.039**	-0.020***	-0.023***	-0.022***
Low correlation simulation	0.005***	-0.012***	-0.015***	-0.005***
High correlation simulation	0.144	0.173	0.145	0.145

Table S.4. Regression analysis. Total SWB.

Estimation of the difference between average happiness with duration weights and average happiness without duration weights using as covariates the within-person correlation between happiness and duration reports, the standard deviation of happiness reports, and the standard deviation of duration reports. Data from EMA reports and the average happiness is computed using the Total formula.

	Student sample $R^2 = 0.831$			**Mixed sample** $R^2 = 0.818$		
	Coefficient	**Std error**	**p-value**	**Coefficient**	**Std error**	**p-value**
Constant	-0.097	0.024	<0.001	-0.026	0.029	0.372
Correlation SWB/duration	0.913	0.028	<0.001	1.075	0.038	<0.001
Standard deviation SWB	0.052	0.014	<0.001	-0.013	0.013	0.346
Standard deviation duration	<0.001	<0.001	0.311	<0.001	<0.001	0.101

Table S.5. Regression analysis. Daily SWB.

Estimation of the difference between average happiness with duration weights and average happiness without duration weights using as covariates the within-person correlation between happiness and duration reports, the standard deviation of happiness reports, and the standard deviation of duration reports. Data from EMA reports and the average happiness is computed using the Daily formula.

	Student sample $R^2 = 0.458$			Mixed sample $R^2 = 0.482$		
	Coefficient	**Std error**	**p-value**	**Coefficient**	**Std error**	**p-value**
Constant	-0.056	0.020	0.006	0.004	0.020	0.840
Correlation SWB/duration	0.315	0.024	<0.001	0.329	0.026	<0.001
Standard deviation SWB	0.027	0.012	0.021	-0.004	0.009	0.690
Standard deviation duration	<0.001	<0.001	0.023	<0.001	<0.001	0.651

Low correlations between intensity and duration remained when grouping overall SWB reports by high or low happiness (Table A.6), and when grouping activities with above or below average SWB (Tables A.7-A.10).

Table S.6. Correlation between SWB and duration by valence.

Significant results marked *** if p-val < 0.001, ** if p-val < 0.01, * if p-val <0.05.

Correlation	Student sample				Mixed sample			
	Happiness		Worthwhileness		Happiness		Worthwhileness	
	EMA	DRM	EMA	DRM	EMA	DRM	EMA	DRM
All episodes	0.01	0.02	0.06***	0.02	-0.01	0.02*	0.01	0.03***
High valence (>7)	0.07***	0.05*	0.08***	-0.01	-0.02	0.01	-0.01	0.01
Low valence (<5)	-0.16***	0.03	-0.10***	0.04	-0.03	0.09*	-0.09**	-0.05

Table S.7. Correlation between SWB and duration for activities with high mean SWB. Significant results marked *** if p-val < 0.001, ** if p-val < 0.01, * if p-val <0.05.

Correlation	Student sample				Mixed sample			
	Happiness		Worthwhileness		Happiness		Worthwhileness	
	EMA	DRM	EMA	DRM	EMA	DRM	EMA	DRM
Exercise	0.08	0.02	-0.04	-0.17**	0.05	0.15*	0.01	-0.07
Conversation	0.19**	0.09	0.14*	0.01	0.17***	-0.07	0.15***	-0.12*
Listening to music	-0.01	-0.10	0.20*	-0.03	0.12**	0.03	0.08	-0.04
Socialising	0.06	0.02	0.03	0.01	NA	NA	NA	NA

Table S.8. Mean SWB for activities with high mean SWB.

Mean SWB	Student sample				Mixed sample			
	Happiness		Worthwhileness		Happiness		Worthwhileness	
	EMA	DRM	EMA	DRM	EMA	DRM	EMA	DRM
Exercise	7.01	7.21	7.05	7.40	7.94	7.74	8.16	8.17
Conversation	7.03	7.21	6.94	7.20	7.62	7.69	7.94	8.00
Listening to music	7.03	7.17	6.92	7.09	7.23	7.46	7.38	7.58
Socialising	7.18	7.50	6.81	7.02	NA	NA	NA	NA

Table S.9. Correlation between SWB and duration for activities with low mean SWB.

Significant results marked *** if p-val < 0.001, ** if p-val < 0.01, * if p-val <0.05.

Correlation	Student sample				Mixed sample			
	Happiness		Worthwhileness		Happiness		Worthwhileness	
	EMA	DRM	EMA	DRM	EMA	DRM	EMA	DRM
Studying	-0.01	0.04	<0.01	-0.04	0.07	0.04	0.07*	0.02
Waiting	-0.14	0.03	0.09	0.09	-0.07	-0.01	-0.01	-0.14
Commuting	0.02	0.04	0.07*	0.03	0.02	0.13**	0.06	0.13**

Table S.10. Mean SWB for activities with low mean SWB.

Mean SWB	Student sample				Mixed sample			
	Happiness		Worthwhileness		Happiness		Worthwhileness	
	EMA	DRM	EMA	DRM	EMA	DRM	EMA	DRM
Studying	6.21	6.21	6.56	6.70	6.78	6.63	7.49	7.42
Waiting	6.13	6.20	5.34	5.74	6.86	6.56	7.06	6.77
Commuting	6.26	6.28	6.11	6.11	6.87	6.80	6.90	7.03

Table S.11. High valence episodes.

Mean of pairwise differences between average SWB with duration weights and average SWB without duration weights for each type of formula, SWB measure, questionnaire type, and sample. In computing Table S.11, we considered only those episodes with SWB greater than 7. Significant results mean that mean difference is less than a threshold of a tenth of a point. (*** if p-val < 0.001, ** if p-val < 0.01, * if p-val < 0.05).

Mean difference	Student sample				Mixed sample			
	EMA		DRM		EMA		DRM	
	Total	Daily	Total	Daily	Total	Daily	Total	Daily
Happiness	0.014***	0.003***	0.038***	0.012***	0.019***	0.007***	0.027***	0.017***
Worthwhileness	0.050***	0.010***	0.011***	0.008***	0.026***	0.009***	0.034***	0.015***

Table S.12. Low valence episodes.

Mean of pairwise differences between average SWB with duration weights and average SWB without duration weights for each type of formula, SWB measure, questionnaire type, and sample. In computing Table S.12, we considered only those episodes with SWB less than 5. Significant results mean that mean difference is less than a threshold of a tenth of a point. (*** if p-val < 0.001, ** if p-val < 0.01, * if p-val < 0.05).

Mean difference	**Student sample**				**Mixed sample**			
	EMA		**DRM**		**EMA**		**DRM**	
	Total	**Daily**	**Total**	**Daily**	**Total**	**Daily**	**Total**	**Daily**
Happiness	0.013***	0.003***	0.008***	-0.004***	-0.031*	0.001***	0.005***	0.019***
Worthwhileness	0.010***	0.004***	0.009***	-0.002***	-0.052	0.008***	0.035**	-0.002***

Table S.13. Mean of pairwise differences between average happiness with duration weights and average worthwhileness without duration weights. Significant results mean that mean difference is less than 1%. (*** if p-val < 0.001, ** if p-val < 0.01).

Mean difference	Student sample				Mixed sample			
	EMA		DRM		EMA		DRM	
	Total averag es	Daily averages	Total averag es	Daily averages	Total averages	Daily averages	Total averages	Daily averages
Worthwhileness	0.151	0.077*	0.017***	0.030***	0.046**	0.028***	0.011***	0.006***

Table S.14. Regression analysis. Happiness v. duration as predictors of current wellbeing report. We ran a regression analysis with current wellbeing (happiness intensity / its duration) as a dependent variable and compared the contribution of the following independent variables: duration report of the previous episode, happiness report of the previous episode. For each individual, we considered each activity that appears at least 3 times in individual's reports. We only looked at previous reports if they occurred on the same day as the current report. Happiness was taken as the maximum score of from people's happiness and worthwhileness reports. Significant results mean that mean difference is less than 1%. (*** if p-val < 0.001, ** if p-val < 0.01).

	Coefficient	**Standard error**	**p value**
Constant	3.690	0.447	3.59e-16***
Happiness from last reported episode (same day)	0.527	0.062	< 2e-16 ***
Duration from last reported episode	-0.659	0.084	8.98e-15***

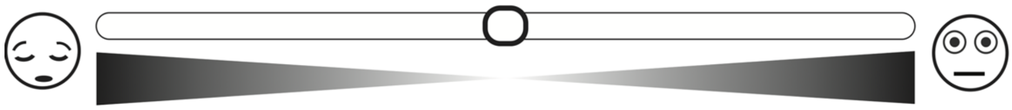
This model shows that the duration of the current as well as the previous episode are strong predictors of the wellbeing variance reported per episode. R2 = 0.07, i.e. this model explains 7% of the variance in the reported wellbeing.

12. Paper 2, Appendix

List of Figures

Figure 1. Affective Slider

Arousal



Valence

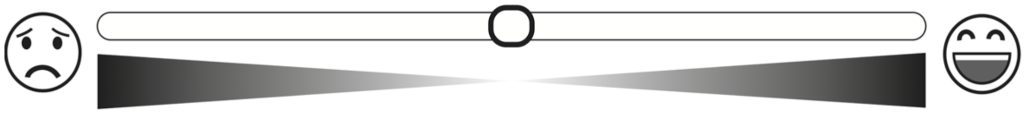
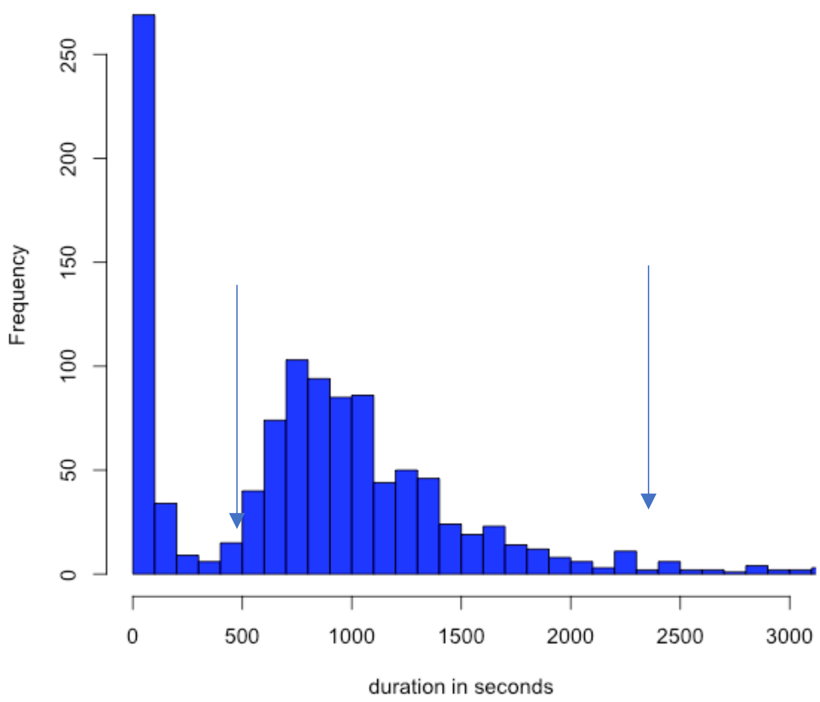
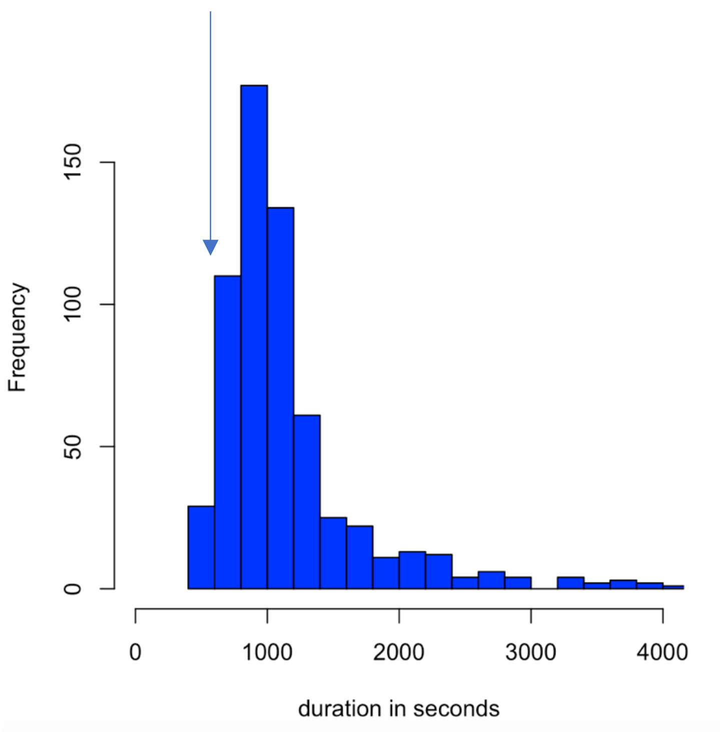


Figure 2. Histogram determining cut-off point for duration-based exclusions study 1



- Those who finished the survey in less than 400 seconds (6.7 minutes) were excluded from the study. This figure was determined using the duration histogram below. The lower tail of the distribution was used as the cut-off point (see left blue arrow).
- Those who finished the survey in more than 2250 seconds (37.5 minutes). The higher tail of the distribution was used as the cut-off point (see right blue arrow). After this point responses became more variable and were therefore considered to be less reliable.

Figure 5: Histogram determining cut-off point for duration-based exclusions study 2



List of Tables

Study 1

Table 1. Number of participants in each gender type

Gender	Count
Male	318
Female	324

Table 2. Number of participants in each age-group

Age group	Count
18-24	40
25-34	94
35-44	126
45-54	149
55-64	120
65+	113

Table 3. Planned comparisons from ANOVA predicting valence change per condition

	diff	lwr	upr	p
Anxious-Neutral	-6.841	-14.849	1.167	0.135
Happy-Neutral	11.561	3.669	19.453	0.001***
Relaxed-Neutral	9.241	1.170	17.312	0.016*
Sad-Neutral	-13.373	-21.545	-5.201	0.000***
Happy-Anxious	18.402	10.541	26.263	0.000***
Relaxed-Anxious	16.081	8.041	24.122	0.000***
Sad-Anxious	-6.532	-14.674	1.610	0.183
Relaxed-Happy	-2.321	-10.245	5.604	0.930
Sad-Happy	-24.934	-32.962	-16.907	0.000***
Sad-Relaxed	-22.613	-30.817	-14.410	0.000***

Table 4. Planned comparisons from ANOVA predicting arousal change per condition

	diff	lwr	upr	p
Anxious-Neutral	-4.542	-12.348	3.263	0.503
Happy-Neutral	4.918	-2.775	12.610	0.405

Relaxed-Neutral	2.702	-5.165	10.569	0.881
Sad-Neutral	-10.576	-18.541	-2.611	0.003**
Happy-Anxious	9.460	1.798	17.122	0.007**
Relaxed-Anxious	7.244	-0.592	15.081	0.086
Sad-Anxious	-6.034	-13.969	1.902	0.230
Relaxed-Happy	-2.216	-9.940	5.508	0.935
Sad-Happy	-15.494	-23.318	-7.670	0.000***
Sad-Relaxed	-13.278	-21.274	-5.282	0.000***

Table 5. Linear regression predicting valence change

	Estimate	SE	Statistic	p
(Intercept)	5.211	8.084	0.645	0.519
Extraversion	0.125	0.390	0.320	0.749
Openness	0.178	0.381	0.467	0.641
Emotional stability	-0.692	0.506	-1.368	0.172
Agreeableness	0.674	0.522	1.289	0.198
Conscientiousness	-0.731	0.488	-1.497	0.135
Wellbeing	0.012	0.196	0.064	0.949
Depression (PHQ8)	-0.188	0.405	-0.464	0.643
Social desirability	0.243	0.422	0.577	0.564
Age	0.199	0.725	0.275	0.784
Gender	-3.750	2.146	-1.748	0.081
R^2	0.012			
F-statistic	0.012			
n	622			

Table 6. Linear regression predicting arousal change

	Estimate	SE	Statistic	p
(Intercept)	-1.752	7.550	-0.232	0.817
Extraversion	-0.283	0.364	-0.778	0.437
Openness	-0.023	0.356	-0.063	0.949
Emotional stability	0.034	0.472	0.071	0.943
Agreeableness	0.837	0.488	1.714	0.087
Conscientiousness	-0.406	0.456	-0.890	0.374
Wellbeing	-0.071	0.183	-0.389	0.697
Depression (PHQ8)	-0.073	0.378	-0.195	0.846
Social desirability	-0.306	0.394	-0.776	0.438
Age	0.563	0.677	0.832	0.406

Gender	1.793	2.004	0.895	0.371
R^2	0.011			
F-statistic	0.714			
n	622			

Table 7. Linear regression on trait anxiety

	Estimate	SE	Statistic	p
(Intercept)	66.022	1.595	41.392	<0.001***
Arousal delta	-0.010	0.011	-0.878	0.380
Valence delta	-0.009	0.010	-0.864	0.388
Agreeableness	-0.035	0.122	-0.286	0.775
Extraversion	-0.304	0.092	-3.291	<0.01*
Emotional stability	-1.093	0.115	-9.49	<0.001***
Openness	0.086	0.090	0.959	0.338
Social desirability	-0.356	0.101	-3.532	0.000
Wellbeing	-0.620	0.047	-13.329	<0.001***
Baseline depression	0.735	0.097	7.608	<0.001***
R^2	0.714			
F-statistic	172.5			
n	623			

Note: anxiety was measured using the Trait component of the State-Trait-Anxiety-Index. *p < 0.05, **p < 0.01, ***p < 0.001

Table 8. Linear regression on trait depression

	Estimate	SE	Statistic	p
(Intercept)	19.957	1.477	13.51	<0.001***
Arousal delta	0.018	0.010	1.78	0.076
Valence delta	0.000	0.009	-0.05	0.964
Agreeableness	0.115	0.113	1.02	0.310
Extraversion	-0.079	0.086	-0.92	0.358
Emotional stability	-0.333	0.107	-3.12	<0.01*
Openness	0.087	0.083	1.04	0.297
Social desirability	-0.392	0.093	-4.19	0.000
Wellbeing	-0.431	0.043	-10.00	<0.001***
Baseline depression	0.729	0.089	8.15	<0.001***
R^2	0.552			
F-statistic	85.13			

n	623

Note: depression was measured using the Becks Depression Inventory. *p < 0.05, **p < 0.01, ***p < 0.001

Study 2

Table 9. Proportion of participants in each gender type

Gender/identity	Count
Male	306
Female	335
Transgender Female	1
Transgender Male	0
Gender variant/non-conforming	1
Not listed	0

Table 10. Proportion of participants in each age-group

Age group	Count
18-24	82
25-34	66
35-44	112
45-54	104
55+	279

Table 11. Proportion of participants in each ethnic group

Ethnicity	Count
Asian / Asian British	56
Black/African / Caribbean / Black British	13
Hispanic or Latino	1
Mixed / Multiple ethnic groups	11
White	557
Other ethnic group	56
Prefer not to say	13

Table 12. Pairwise comparisons for ANOVA on mood predicting valence delta

	diff	lwr	upr	p
Happy-Anxious	21.516	12.891	30.141	0.000***
Neutral-Anxious	13.164	4.702	21.626	0.000***
Relaxed-Anxious	18.660	9.934	27.386	0.000***
Sad-Anxious	-9.707	-18.332	-1.082	0.018*
Neutral-Happy	-8.352	-16.834	0.130	0.056
Relaxed-Happy	-2.856	-11.600	5.889	0.899
Sad-Happy	-31.223	-39.867	-22.579	0.000***
Relaxed-Neutral	5.496	-3.088	14.080	0.403
Sad-Neutral	-22.871	-31.353	-14.390	0.000***
Sad-Relaxed	-28.367	-37.112	-19.623	0.000***

Table 13. Pairwise comparisons for ANOVA on mood predicting arousal delta

	diff	lwr	upr	p
Happy-Anxious	12.040	2.909	21.171	0.003**
Neutral-Anxious	7.579	-1.379	16.538	0.141
Relaxed-Anxious	9.196	-0.042	18.433	0.052
Sad-Anxious	-4.076	-13.207	5.055	0.739
Neutral-Happy	-4.461	-13.440	4.519	0.654
Relaxed-Happy	-2.845	-12.102	6.413	0.918
Sad-Happy	-16.116	-25.267	-6.965	0.000***
Relaxed-Neutral	1.616	-7.472	10.704	0.989
Sad-Neutral	-11.655	-20.635	-2.676	0.004**
Sad-Relaxed	-13.272	-22.529	-4.014	0.001**

Table 14. Linear regression predicting valence change

	Estimate	SE	Statistic	p
(Intercept)	-13.916	7.722	-1.802	0.072
Extraversion	0.262	0.429	0.610	0.542
Openness	0.071	0.412	0.173	0.863
Emotional stability	-0.092	0.561	-0.164	0.870
Agreeableness	-0.389	0.623	-0.625	0.532
Conscientiousness	0.770	0.558	1.381	0.168
Wellbeing	-0.131	0.183	-0.717	0.474

Social desirability	0.674	0.437	1.544	0.123
Age	0.559	0.866	0.646	0.519
Gender	0.658	2.109	0.312	0.755
R^2	0.013			
F-statistic	0.807			
n	555			

Table 15. Linear regression predicting arousal change

	Estimate	SE	Statistic	p
(Intercept)	-5.001	7.516	-0.665	0.506
Extraversion	-0.702	0.418	-1.681	0.093
Openness	0.352	0.401	0.879	0.380
Emotional stability	0.443	0.546	0.811	0.418
Agreeableness	-0.577	0.606	-0.951	0.342
Conscientiousness	-0.016	0.543	-0.029	0.977
Wellbeing	0.229	0.179	1.285	0.200
Social desirability	0.120	0.425	0.281	0.779
Age	0.408	0.843	0.485	0.628
Gender	3.175	2.053	1.547	0.123
R^2	0.015			
F-statistic	0.921			
n	555			

Table 16. Positive word evaluation score by group

Group	Count	Mean positive word evaluation	Standard deviation
Anxious	113	6.02	1.83
Happy	112	6.29	1.85
Neutral	121	5.87	1.84
Relaxed	107	5.87	1.84
Sad	112	6.09	1.79

Table 17. Negative word evaluation score by group

Group	Count	Mean negative word evaluation	Standard deviation
Anxious	113	2.78	1.82
Happy	112	3.18	1.83
Neutral	121	3.15	2.08
Relaxed	107	2.65	1.62
Sad	112	3.21	2.04

Table 18. Simple linear regression model on trait anxiety (study 2)

	Estimate	SE	Statistic	p
(Intercept)	78.551	1.421	55.273	<0.001***
Arousal delta	-0.003	0.011	-0.271	0.786
Valence delta	-0.010	0.011	-0.923	0.357
Agreeableness	0.350	0.146	2.397	<0.05*
Extraversion	-0.420	0.103	-4.093	<0.001***
Emotional stability	-1.696	0.123	-13.747	<0.001***
Openness	-0.243	0.097	-2.513	<0.05*
Social desirability	-0.678	0.103	-6.59	<0.001***
Wellbeing	-0.688	0.043	-15.854	<0.01**
(Intercept)	78.551	1.421	55.273	<0.001***
R^2	0.760			
F-statistic	220.6			
n	556			

Note: anxiety was measured using the Trait component of the State-Trait-Anxiety-Index. *p < 0.05, **p < 0.01, ***p < 0.001

Table 19. Simple linear regression model on trait depression (study 2)

	Estimate	SE	Statistic	p
(Intercept)	32.189	1.546	20.814	<0.001***
Arousal delta	-0.031	0.012	-2.532	0.011*
Valence delta	0.022	0.012	1.795	0.073
Agreeableness	0.150	0.159	0.941	0.347
Extraversion	-0.117	0.112	-1.043	0.297
Emotional stability	-0.774	0.134	-5.765	<0.001***
Openness	0.026	0.105	0.243	0.808
Social desirability	-0.297	0.112	-2.654	<0.01**

Wellbeing	-0.658	0.047	-13.94	<0.001***
(Intercept)	32.189	1.546	20.814	<0.001***
R^2	0.557			
F-statistic	87.49			
n	556			

Note: depression was measured using the Becks Depression Inventory. *p < 0.05, **p < 0.01, ***p < 0.001

Table 20. Tukey adjusted pairwise comparisons from the significant ANOVA on state depression

	diff	lwr	upr	p adj
Happy-Anxious	-2.731	-5.183	-0.278	0.020*
Neutral-Anxious	-1.384	-3.790	1.023	0.515
Relaxed-Anxious	-1.224	-3.705	1.258	0.660
Sad-Anxious	-0.445	-2.898	2.008	0.988
Neutral-Happy	1.347	-1.065	3.759	0.544
Relaxed-Happy	1.507	-0.980	3.994	0.461
Sad-Happy	2.286	-0.172	4.744	0.082
Relaxed-Neutral	0.160	-2.281	2.601	1.000
Sad-Neutral	0.939	-1.473	3.351	0.824
Sad-Relaxed	0.779	-1.708	3.266	0.912

Table 21. Simple linear regression model on state anxiety (study 2)

	Estimate	SE	Statistic	p
(Intercept)	30.600	1.171	26.137	0.000***
Arousal delta	-0.010	0.009	-1.031	0.303
Valence delta	-0.002	0.009	-0.201	0.841
Agreeableness	0.173	0.120	1.44	0.151
Extraversion	0.062	0.085	0.735	0.463
Emotional stability	-0.743	0.102	-7.307	0.000***
Openness	0.107	0.080	1.335	0.182
Social desirability	-0.160	0.085	-1.89	0.059
Wellbeing	-0.382	0.036	-10.692	0.000***
R^2	0.473			

F-statistic	62.26
n	556

Note: depression was measured using the anxiety component of the State-Trait-Anxiety-Depression-Index.
*p < 0.05, **p < 0.01, ***p < 0.001

Table 22. Simple linear regression model on state depression (study 2)

	Estimate	SE	Statistic	p
(Intercept)	42.522	0.936	45.406	0.000***
Arousal delta	-0.015	0.007	-2.041	0.045*
Valence delta	-0.020	0.007	-2.777	0.005**
Agreeableness	-0.150	0.097	-1.549	0.12198
Extraversion	-0.313	0.068	-4.581	0.000***
Emotional stability	-0.586	0.076	-7.729	0.000***
Openness	0.018	0.064	0.281	0.77902
Social desirability	-0.128	0.067	-1.904	0.05744
Wellbeing	-0.759	0.049	-15.521	0.000***
(Intercept)	42.522	0.936	45.406	0.000***
R^2	0.630			
F-statistic	118.1			
n	556			

Note: depression was measured using the depression component of the State-Trait-Anxiety-Depression-Index. *p < 0.05, **p < 0.01, ***p < 0.001

13. Paper 3, Appendix

List of tables

A.1. Kendal correlations between all social behaviours were (mostly) significantly associated with each other but small in magnitude

	Social distancing	Self-isolating	Avoiding crowds	Avoiding small groups	Less in person interaction	Less overall interaction
Social distancing						
Self-isolating	0.184***					
Avoiding crowds	0.553***	0.176***				
Avoiding small groups	0.454***	0.273***	0.489***			
Less in person interaction	0.323***	0.0103**	0.357***			
Less overall interaction	0.177***	0.0150***	0.148***	0.183***	0.410***	
Mask outdoors	-0.025	0.161***	-0.052	0.039	0.01	0.075*

A.2. Social behaviours 1-7 means and standard deviations across waves

	Social distancing	Self-isolating	Avoiding crowds	Avoiding small groups	Less in person interaction	Less overall interaction	Mask outdoors
Time 1	M=9.098 SD=1.481	M=4.391 SD=4.046	M=9.356 SD=1.290	M=8.603 SD=2.224	M=8.592 SD=2.438	M=7.023 SD=3.149	M=2.397 SD=3.590
Time 2	M=8.661 SD=1.826	M=3.598 SD=3.860	M=9.023 SD=1.703	M=7.672 SD=2.907	M=8.080 SD=2.646	M=6.655 SD=3.151	M=3.534 SD=3.952
Time 3	M=8.529 SD=1.685	M=3.178 SD=3.842	M=8.598 SD=2.112	M=6.793 SD=3.285	M=7.799 SD=2.803	M=6.483 SD=3.147	M=5.218 SD=3.968

Happy tendency descriptive statistics and associated panel linear model

A.3. Happy tendency means across waves

Wave	Mean	SD	Count
1	23.9	357	175
2	49.6	412	175
3	48.4	317	175

Happy tendency is relatively low in wave 1 and almost doubles in waves 2 and 3. However, a repeated measures ANOVA with time as the independent variable and happy tendency as the dependent variable confirmed that these changes were not significant (F(2,336) = 0.289, p = 0.741).

A.4. Simple panel linear model on happy tendency: fixed effects

	Estimate	Std. Error	p
Social distancing	3.677	15.680	0.815
Self-isolating	-10.860	7.829	0.166
Avoiding crowds	-19.764	16.477	0.231
Avoiding small groups	15.818	9.352	0.092
Less in person interaction	-15.463	10.532	0.143
Less overall interaction	4.757	7.915	0.548
Mask outdoors	4.730	7.507	0.529
N	175		
F	1.177		
R^2	0.024		

Notes: Fixed effects regression using prescribed social behaviours (independent variables) to predict happy tendency (dependent variable). Standard errors are clustered on an individual level. * p<0.05, ** p<0.01, *** p<0.001

A.5. Simple panel linear models 1-7 on happy tendency, each model focusing on a single social behaviour in isolation: fixed effects

	Dependent variable:						
	Happy tendency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social distancing	-2.773						
	(14.604)						
Self-isolating		-10.463					
		(7.615)					
Avoiding crowds			-18.188				
			(14.873)				
Avoiding small groups				8.995			
				(8.757)			
Less in person interaction					-12.272		
					(9.182)		
Less overall interaction						-0.003	
						(7.018)	
Mask outdoors							6.770
							(7.443)
Observations	525	525	525	525	525	525	525
R^2	0.000	0.005	0.004	0.003	0.005	0.000	0.002
Adjusted R^2	-0.510	-0.502	-0.504	-0.506	-0.502	-0.510	-0.506
F Statistic (df = 1; 347)	0.036	1.888	1.495	1.055	1.786	0.000	0.827

Notes: Fixed effects regression using each prescribed social behaviour (independent variable) to predict happy tendency (dependent variable) in separate models. Standard errors are clustered on an individual level. * p<0.05, ** p<0.01, *** p<0.001

Social tendency descriptive statistics and associated panel linear model

A.6. Social tendency means across waves

Wave	Mean	SD	Count
1	53.5	339	175
2	31.3	323	175
3	19.6	354	175

Social tendency is relatively high in wave 1 and reduces in waves 2 and 3. However, a repeated measures ANOVA with time as the independent variable and social tendency as the dependent variable confirmed that these changes were not significant (F(2,342) = 0.48, p = 0.619).

A.7. Simple panel linear model on social tendency: fixed effects

	Estimate	**Std. Error**	**p**
Social distancing	-12.8201	14.4748	0.3764
Self-isolating	4.0108	7.3191	0.5841
Avoiding crowds	-5.6714	15.2022	0.7093
Avoiding small groups	-7.7185	8.6102	0.3707
Less in person interaction	14.3128	9.7198	0.1418
Less overall interaction	-3.4199	7.4078	0.6446
Mask outdoors	1.3941	6.9156	0.8404
N	175		
F	0.604		
R^2	0.012		

Notes: Fixed effects regression using prescribed social behaviours (independent variables) to predict social tendency (dependent variable). Standard errors are clustered on an individual level.
* p<0.05, ** p<0.01, *** p<0.001

A.8. Simple panel linear models 1-7 on social tendency, each model focusing on a single social behaviour in isolation: fixed effects

	Dependent variable:						
	Social tendency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social distancing	-14.581						
	(13.344)						
Self-isolating		2.546					
		(7.094)					
Avoiding crowds			-9.219				
			(13.640)				
Avoiding small groups				-5.875			
				(8.026)			
Less in person interaction					9.872		
					(8.389)		

Less overall interaction						-0.025	
						(6.478)	
Mask outdoors							0.472
							(6.823)
Observations	516	516	516	516	516	516	516
R^2	0.003	0.000	0.001	0.002	0.004	0.000	0.000
Adjusted R^2	-0.505	-0.510	-0.508	-0.508	-0.504	-0.510	-0.510
F Statistic (df = 1; 341)	1.194	0.129	0.457	0.536	1.385	0.000	0.005

Notes: Fixed effects regression using each prescribed social behaviour (independent variable) to predict social tendency (dependent variable) in separate models. Standard errors are clustered on an individual level. * p<0.05, ** p<0.01, *** p<0.001

Robustness checks for the association between less overall interaction and sad tendency

A.9. Simple panel linear models 1-7 on sad tendency, each model focusing on a single social behaviour in isolation: fixed effects

	Dependent variable:						
	Sad tendency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social distancing	5.126						
	(11.906)						
Self-isolating		-9.759					
		(6.184)					
Avoiding crowds			7.574				
			(12.116)				
Avoiding small groups				-5.575			
				(7.143)			
Less in person interaction					-10.833		
					(7.467)		
Less overall interaction						-17.517***	
						(5.643)	
Mask outdoors							1.796
							(6.056)
Observations	522	522	522	522	522	522	522
R^2	0.001	0.007	0.001	0.002	0.006	0.027	0.000
Adjusted R^2	-0.509	-0.499	-0.508	-0.507	-0.501	-0.469	-0.510

F Statistic (df = 1; 345)	0.185	2.490	0.391	0.609	2.105	9.637^{***}	0.088

Notes: Fixed effects regression using each prescribed social behaviour (independent variable) to predict social tendency (dependent variable) in separate models. Standard errors are clustered on an individual level. * p<0.05, ** p<0.01, *** p<0.001

A.10. Simple panel linear model on sad tendency: random effects

	Estimate	**Std. Error**	**p**
(Intercept)	134.596	87.473	0.124
Social distancing	1.490	9.819	0.879
Self-isolating	-2.281	3.881	0.557
Avoiding crowds	-3.465	9.918	0.727
Avoiding small groups	-0.269	5.649	0.962
Less in person interaction	-4.429	6.141	0.471
Less overall interaction	-10.244	4.934	0.038*
Mask outdoors	-2.795	3.544	0.430
N	174		
Chisq	11.167		
R^2	0.021		
Cohen's *d*	0.157		

Notes: Random effects regression using prescribed social behaviours (independent variables) to predict sad tendency (dependent variable). Standard errors are clustered on an individual level. Cohen's *d* value was caculated using the behaviour 6 coefficient to estimate effect size. * p<0.05, ** p<0.01, *** p<0.001

A.11. Covariate panel linear model on sad tendency **including** time non-varying variables: random effects

	Estimate	Std. Error	z-value	p
(Intercept)	1537.403	390.062	3.941	0.000***
Less overall interaction	-17.521	2.643	-6.628	0.000***
Corona fear	2.355	4.500	0.523	0.601
Anxious	6.768	5.437	1.245	0.213
Happy	-14.872	6.657	-2.234	0.025*
Bis	0.413	3.102	0.133	0.894
Bas Drive	8.942	4.137	2.162	0.031*
Bas Reward	-0.466	1.809	-0.258	0.797
Hours away from home	5.712	6.963	0.820	0.412
Stressed	-19.912	5.568	-3.576	0.000***

Social Anxiety	-0.166	0.409	-0.406	0.685
Anxiety	-6.582	1.641	-4.011	0.000***
Mental Health	-12.879	19.362	-0.665	0.506
SWB	-11.722	3.636	-3.224	0.001**
Extraversion	-1.315	1.459	-0.901	0.368
Conscientiousness	5.926	7.191	0.824	0.410
Stability	-13.774	8.610	-1.600	0.110
Agreeableness	5.010	7.873	0.636	0.525
Openness	-7.207	6.620	-1.089	0.276
Age	-33.277	13.479	-2.469	0.014*
Gender	-81.893	35.112	-2.332	0.020*
Ethnicity (prefer not to say)	-223.477	221.507	-1.009	0.313
White	-125.585	198.824	-0.632	0.528
Asian	-114.670	250.763	-0.457	0.647
Asian/Black	-261.064	294.797	-0.886	0.376
Income	6.125	7.238	0.846	0.397
Region	-2.451	4.495	-0.545	0.586
Cardiovascular disease	-166.668	88.626	-1.881	0.060
Diabetes	-41.742	70.208	-0.595	0.552
Chronic respiratory disease	-20.056	102.294	-0.196	0.845
Hypertension	30.450	67.805	0.449	0.653
Asthma	-13.148	67.888	-0.194	0.846
Other serious condition	-9.602	75.259	-0.128	0.898
No health condition	-8.253	72.309	-0.114	0.909
Children	2.219	14.930	0.149	0.882
House living	-1.033	11.384	-0.091	0.928
Key worker	-49.456	40.239	-1.229	0.219
Covid symptoms (none)	-175.356	199.242	-0.880	0.379
Covid Symptoms (s.o I know living apart)	-251.760	205.949	-1.222	0.222
Covid Symptoms (myself)	-88.939	226.682	-0.392	0.695
Covid Symptoms (myself + s.o I live with)	-338.779	225.068	-1.505	0.132
Covid Symptoms (s.o I know)	-203.559	280.723	-0.725	0.468
Covid Symptoms (s.o I live with)	-173.106	247.914	-0.698	0.485
N	174			
Chisq	112.736			
R^2	0.070			
Cohen's *d* for less overall interaction	-0.503			

Notes: Random effects panel linear regression using less overall interaction and all other time varying *and* non time varying covariates (independent variables) to predict sad tendency (dependent variable). Cohen's *d* value was caculated using the less overall interaction coefficient to estimate effect size. * p<0.05, ** p<0.01, *** p<0.001

Testing for associations between affective variables and reduced-overall-interaction

A.12. Testing for associations between trait level affective variables with reduced-overall-interaction, fixed effects panel linear model

	Dependent variable:		
	Reduced-overall-interaction		
	(1)	(2)	(3)
Anxiety	0.009		
	(0.035)		
Social Anxiety		0.003	
		(0.009)	
Life Satisfaction			-0.028
			(0.184)
Observations	522	522	522
R^2	0.0002	0.0004	0.0001
Adjusted R^2	-0.510	-0.510	-0.510
F Statistic (df = 1; 345)	0.068	0.132	0.024
Note:	*p<0.1; **p<0.05; ***p<0.01		

A.13. Testing for associations between coronavirus specific affective variables with reduced-overall-interaction, fixed effects panel linear models

	Dependent variable:	
	Reduced-overall-interaction	
	(1)	(2)
Corona fear (self)	0.155	
	(0.121)	
Corona fear (others)		0.115
		(0.113)
Observations	522	522
R^2	0.005	0.003
Adjusted R^2	-0.503	-0.506
F Statistic (df = 1; 345)	1.632	1.031
Note:	*p<0.1; **p<0.05; ***p<0.01	

A.14. Testing for associations between state level affective variables with reduced-overall-interaction, fixed effects panel linear model

	Dependent variable:				
	Reduced-overall-interaction				
	(1)	(2)	(3)	(4)	(5)
Stressed	-0.045				
	(0.091)				
Anxious		0.142			
		(0.089)			
Happy			-0.076		
			(0.129)		
Valence				-0.009	
				(0.010)	
Arousal					-0.001
					(0.008)
Observations	522	522	522	522	522
R^2	0.001	0.007	0.001	0.002	0.00001
Adjusted R^2	-0.509	-0.499	-0.509	-0.507	-0.510
F Statistic (df = 1; 345)	0.240	2.536	0.348	0.787	0.005
Note:		*p<0.1; **p<0.05; ***p<0.01			

Testing for associations between affective variables and sad tendency

A.15. Testing for associations between trait level affective variables with sad tendency, fixed effects panel linear model

	Dependent variable:		
	Sad tendency		
	(1)	(2)	(3)
Anxiety	-8.703**		
	(3.667)		
Social Anxiety		-0.621	
		(0.993)	
Life Satisfaction			12.193
			(19.545)
Observations	522	522	522
R^2	0.016	0.001	0.001
Adjusted R^2	-0.486	-0.508	-0.508
F Statistic (df = 1; 345)	5.633**	0.391	0.389

Note: *p<0.1; **p<0.05; ***p<0.01

A.16. Testing for associations between coronavirus specific affective variables with sad tendency, fixed effects panel linear model

	Dependent variable:	
	Sad tendency	
	(1)	(2)
Corona fear (self)	3.322	
	(12.893)	
Corona fear (other)		-3.914
		(12.020)
Observations	522	522
R^2	0.0002	0.0003
Adjusted R^2	-0.510	-0.510
F Statistic (df = 1; 345)	0.066	0.106
Note:	*p<0.1; **p<0.05; ***p<0.01	

A.17. Testing for associations between state level affective variables with sad tendency, fixed effects panel linear model

	Dependent variable:				
	Sad tendency				
	(1)	(2)	(3)	(4)	(5)
Stressed	-19.030**				
	(9.635)				
Anxious		-9.758			
		(9.494)			
Happy			-6.046		
			(13.666)		
Valence				0.745	
				(1.060)	
Arousal					-1.009
					(0.859)
Observations	522	522	522	522	522
R^2	0.011	0.003	0.001	0.001	0.004
Adjusted R^2	-0.493	-0.506	-0.509	-0.508	-0.504
F Statistic (df = 1; 345)	3.901**	1.056	0.196	0.494	1.379
Note:			*p<0.1; **p<0.05; ***p<0.01		

Full breakdown of participant characteristics

A.18. Proportion of participants in each age-group

Age group	Count
18-24	2%
25-34	15%
35-44	14%
45-54	19%
55-64	20%
65+	30%

A.19. Proportion of participants in each physical health group

(1 = Excellent, 5 – Poor)

Physical Health	Count
1	10%
2	26%
3	39%
4	18%
5	7%

A.20. Proportion of participants in each income group

Income brackets	Wave 1
Under £5000	6%
£5000-£9999	9%
£10,000-£14,999	10%
£15,000-£19,999	9%
£20,000-£24,000	13%
25,000-£34,999	23%
£35,000-£44,999	11%
£45,000-£54,999	9%
£55,000-£99,999	9%
£100,000+	3%

Income split followed a normal distribution across income brackets with less people at the lower and higher bounds, as expected (this information was only collected in wave 1)

A.21. Proportion of participants in each UK region

	Count
Greater London	15%
South East	16%
South West	8%
West Midlands	10%
North West	8%
North East	6%
Yorkshire and the Humber	5%
East Midlands	8%
East Anglia	9%
Scotland	9%
Northern Ireland	1%
Wales	4%

Comparisons between final sample participants and dropouts

A.22. A comparison of gender in final sample with dropouts

Full sample

Gender	Count
Male	53%
Female	47%

Dropouts

Gender	Count
Male	47%
Female	53%

A.23. A comparison of age in final sample with dropouts

Full sample

Age group	Count
18-24	2%
25-34	15%
35-44	14%
45-54	19%
55-64	20%
65+	30%

Dropouts

Age group	Count
18-24	3%
25-34	14%
35-44	17%
45-54	21%
55-64	17%
65+	27%

A.24. A comparison of income in final sample with dropouts

Full sample

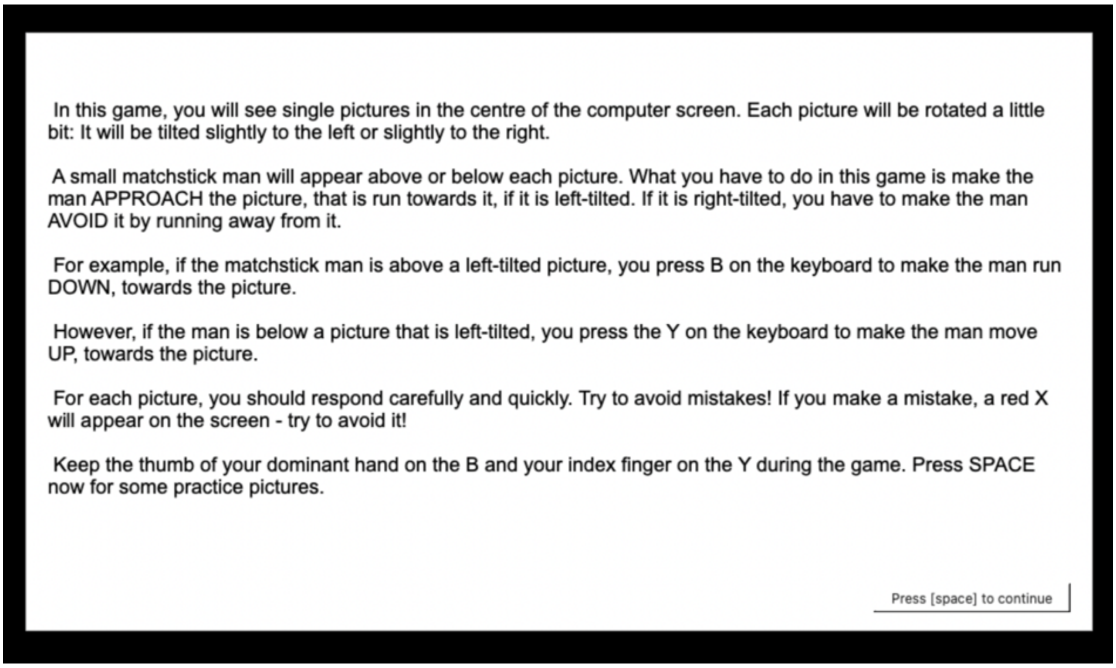
Income brackets	Wave 1
Under £5000	6%
£5000-£9999	9%
£10,000-£14,999	10%
£15,000-£19,999	9%
£20,000-£24,000	13%
25,000-£34,999	23%
£35,000-£44,999	11%
£45,000-£54,999	9%
£55,000-£99,999	9%
£100,000+	3%

Dropouts

Income brackets	Wave 1
Under £5000	8%
£5000-£9999	8%
£10,000-£14,999	13%
£15,000-£19,999	10%
£20,000-£24,000	10%
25,000-£34,999	13%
£35,000-£44,999	9%
£45,000-£54,999	6%
£55,000-£99,999	9%
£100,000+	3%
NA	10%

List of figures

A.1. SCRT instructions



Sample size calculation

Considering that we could not identify appropriate statistical packages that could calculate power for the exact panel linear models we aimed to use, we computed several power analyses for Person correlations, given that they are broadly comparable to our panel models. In line with current practices in the field of psychology, we first calculated the sample size needed to detect a medium effect size (r = .30), assuming the alpha level of 0.05 and the power level of 0.80 (Faul et al., 2009). The analysis indicated that 82 participants would need to be tested. We then ran the same analysis but changed the alpha level to a more conservative 0.001. In this case, the sample size requirement was 179. Overall, this analysis indicated that testing roughly 179 participants would be sufficient to detect medium effects, regardless of whether a standard (0.05) or more conservative (0.001) alpha level is used. Given that we eventually obtained a comparable sample size (i.e., 174), we then computed sensitivity power analyses (Faul et al., 2009) to identify the smallest effect size that could be reliable obtained with the sample size we recruited, assuming the power of 0.80. These analyses showed that the study is likely to be sufficiently powered to detect effect size r = 0.21 (assuming the alpha level of 0.05) and effect size r = 0.30 (assuming the alpha level of 0.001). Therefore, it is plausible that the present research was well powered to detect at least medium effect sizes, regardless of the alpha level used.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Data access

Data files for this this study and associated coding guidelines can be found here https://osf.io/ydqgm/

14. Paper 4, Appendix

A1. ESM questionnaire (study 1 and 2)

1) During the past few hours, I was:

- Working
- Commuting
- Eating
- Shopping
- At the supermarket
- Emails
- Listening to Music
- Taking a nap
- Waiting
- Reading
- Exercise (gym, jogging, swimming, add more, etc.)
- Studying
- Taking a shower
- Conversation
- Watching TV
- Using social media
- Browsing internet
- Playing video games
- ADD [Open text box]

1a) Custom Student Activity List

- Commuting
- Eating
- Exercising
- Socialising

- Watching TV
- Social Media
- Lecture
- Seminar
- Exam
- Studying
- Meeting
- Careers centre
- IT support
- Counselling
- Human resources
- Society/club
- ADD [Open text box]

1b) Custom Staff Activity List

- Commuting
- Eating
- Exercising
- Socialising
- Watching TV
- Teaching
- Presentation
- Admin
- Meeting
- Emails
- IT support
- Counselling
- Human resources
- Finance
- Overtime
- Cleaning
- ADD NEW [Open text box]

2) How long have you been doing this?

Drop down in 10-minute increments, from 10 minutes to 4hr 10min.

3) **I was with:**

- Colleagues
- Friend(s)
- Kids
- Partner
- People that I didn't know before
- Alone
- Family
- ADD [Open text box]

4) I was thinking about: [TREATMENT GROUP ONLY]

- Current activity
- Kids
- Health
- Partner
- Friends
- Events from last day(s)
- Events from my past
- Food
- About tomorrow
- About my future
- ADD [Open text box]

4a) [AFTER SELECTING A THOUGHT] **This thought was: [TREATMENT GROUP ONLY]**

- Positive
- Neutral
- Negative

5) Where are you?

- Home

- Work
- University
- Library
- Sports facility
- Public transport
- Restaurant
- Bar/club
- Cinema
- Supermarket
- Street/outdoors
- At my parents' house
- At my friends house
- Holidays, away from my home city
- ADD [Open text box]

6) How happy did you feel? [TREATMENT GROUP ONLY]

SCALE: Not at all 0 1 2 3 4 5 6 7 8 9 10 Completely

7) How worthwhile did this feel? [TREATMENT GROUP ONLY]

SCALE: Not at all 0 1 2 3 4 5 6 7 8 9 10 Completely

8) How do you feel at the moment? (Choose the one that describes your current state best) **[TREATMENT GROUP ONLY]**

+ Additional info from the final EMA of the day [TREATMENT GROUP ONLY]

1. Overall, how satisfied were you with your life **today**? **[TREATMENT GROUP ONLY]**
2. Overall, to what extent do you feel the things you did **today** were worthwhile? **[TREATMENT GROUP ONLY]**
3. Overall, how happy did you feel **today**? **[TREATMENT GROUP ONLY]**
4. Overall, how anxious did you feel **today**? **[TREATMENT GROUP ONLY]**
5. How many hours did you sleep last night?

SCALE: <6, 6 - 7, 7 - 8, 8 - 9, 9 -10, 10 - 11, >11

6. How much time did you spend doing sport today?

SCALE: 0 - 4 hours

7. How productive were you today?

SCALE: Completely unproductive 0 1 2 3 4 5 6 7 8 9 10 Very productive

8. How much quality time did you have with friends today?

SCALE: No quality time at all 0 1 2 3 4 5 6 7 8 9 10 Lots of quality time

A2. DRM questionnaire (all studies)

Instructions: Think of your day yesterday as a continuous series of scenes or episodes in a film. Give each episode a brief name that will help you remember it (for example, ‘commuting to work', or ‘at lunch with B'...). Write down the approximate times at which each episode began and ended.

Start time - End Time

__:__ until __:__

I was doing:

- Working
- Commuting
- Eating
- Shopping
- At the supermarket
- Emails
- Listening to Music
- Taking a nap
- Waiting
- Reading
- Exercise (gym, jogging, swimming, add more, etc.)
- Studying

- Taking a shower
- Conversation
- Watching TV
- Using social media
- Browsing internet
- Playing video games
- ADD [Open text box]

1a) Custom Student Activity List

- Commuting
- Eating
- Exercising
- Socialising
- Watching TV
- Social Media
- Lecture
- Seminar
- Exam
- Studying
- Meeting
- Careers centre
- IT support
- Counselling
- Human resources
- Society/club
- ADD [Open text box]

1b) Custom Staff Activity List

- Commuting
- Eating
- Exercising

- Socialising
- Watching TV
- Teaching
- Presentation
- Admin
- Meeting
- Emails
- IT support
- Counselling
- Human resources
- Finance
- Overtime
- Cleaning
- ADD NEW [Open text box]

2) I was with:

- Colleagues
- Friend(s)
- Kids
- Partner
- People that I didn't know before
- Alone
- Family
- ADD [Open text box]

3) I was thinking about: [TREATMENT GROUP ONLY]

- Current activity
- Kids
- Health
- Partner
- Friends
- Events from last day(s)

- Events from my past
- Food
- About tomorrow
- About my future
- ADD [Open text box]

3a) [AFTER SELECTING A THOUGHT] **This thought was: [TREATMENT GROUP ONLY]**

- Positive
- Neutral
- Negative

4) Where are you?

- Home
- Work
- University
- Library
- Sports facility
- Public transport
- Restaurant
- Bar/club
- Cinema
- Supermarket
- Street/outdoors
- At my parents' house
- At my friends house
- Holidays, away from my home city
- ADD [Open text box]

5) How happy did you feel? [TREATMENT GROUP ONLY]

SCALE: Not at all 0 1 2 3 4 5 6 7 8 9 10 Completely

6) How worthwhile did this feel? [TREATMENT GROUP ONLY]

SCALE: Not at all 0 1 2 3 4 5 6 7 8 9 10 Completely

A3. Onboarding questionnaire (study 1 and 2)

A. ONS-4

1. Overall, how satisfied are you with your life nowadays?
2. Overall, to what extent do you feel the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?
4. Overall, how anxious did you feel yesterday?
5. Overall, how happy do you feel in general?
6. Overall, how anxious do you feel in general?

B. Happy/worthwhile balance

1. What is your balance between happy and worthwhile experiences nowadays? Move the slider below until it reaches a point that best represents your balance, then click the next button down there on the right.

0. Happy
1. Worthwhile

2. And what is your ideal balance between happy and worthwhile experiences nowadays? Move the slider below until it reaches a point that best represents your ideal balance, then click the next button down there on the right.

0. Happy
1. Worthwhile

C. Basic demographic information (Students and staff)

- Age
- Gender
- Employment status
- Part- or full-time work
- Monthly income (brackets)
- Social media usage (Facebook, Instagram)

+ In study 2 (LSE sample)

- Staff or student status
- Primary and secondary courses
- Year of study
- Study level
- Overseas or national
- Residence (halls or home)
- Funding status
- How would you rate the sense of community at the LSE? (scale of 1-10)
- How well do you feel you fit in with the community at the LSE? (scale of 1-10)
- How much pressure do you feel from the following sources during your academic life? (scale of 1-10)
    - Work load
    - Social life
    - Faculty
    - Family
    - Peers

D. Individual specific characteristic questionnaires

- Self-esteem
- Delayed gratification
- Sense of control
- Optimism
- Big-5 Personality
- Attitude towards uncertainty

A4. Weekly questions (study 1 and 2) **[TREATMENT GROUP ONLY]**

Note that the following questions refer to the past week!

1. Overall, how satisfied were you with your life **this week**?
2. Overall, to what extent do you feel the things you did **this week** were worthwhile?
3. Overall, how happy did you feel **this week**?

4. Overall, how anxious did you feel **this week**?
5. Select the days on which you expect to have events that you are looking forward to (during the following week). (type=checkbox)
   - Monday
   - Tuesday
   - Wednesday
   - Thursday
   - Friday
   - Saturday
   - Sunday
   - None
6. Select the days on which you expect to have events that you are NOT looking forward to (during the following week)..
   - Monday
   - Tuesday
   - Wednesday
   - Thursday
   - Friday
   - Saturday
   - Sunday
   - None

A5. Exit questionnaire (study 1 and 2)

Title: **"Exit questionnaire"**

1. Overall, how satisfied are you with your life nowadays?
2. Overall, to what extent do you feel the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?
4. Overall, how anxious did you feel yesterday?
5. Overall, how happy do you feel in general?
6. Overall, how anxious do you feel in general?

7. Do you feel that you are more aware of what drives your mood now than you were at the beginning of the study? (Likert scale: 5-much more, 4-a little more, 3-no change, 2-a little less, 1-much less)
8. Have you changed your perception on any of the activities, thoughts or people listed during this study? (Likert scale: 5-much more, 4-a little more, 3-no change, 2-a little less, 1-much less)

A6. Onboarding and exit questionnaire (study 3)

A. Demographic information [ONBOARDING ONLY]

- Age
- Gender
- Employment status
- Part- or full-time work
- Monthly income (brackets)
- Student status
- Year of study
- Study level
- Region of origin
- Place of residence during Covid-19
- Current place of residence
- Funding status

B. ONS 4 extended

1. Overall, how satisfied are you with your life?
2. Overall, to what extent do you feel the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?

4. Overall, how anxious did you feel yesterday?

5. Overall, how happy do you feel in general?

6. Overall, how anxious do you feel in general?

7. How happy do you feel today?

Note that some questions that you answered refer to different kinds of being happy. Imagine that someone could ask you only one of the following two questions for the rest of your life to understand your happiness. Which one would you consider as the most important question:
[ ] Overall, how satisfied are you with your life?
[ ] How happy do you feel today? (repeatedly asked)

C. PSS-10

Introduction: The questions in this scale ask about your feelings and thoughts during the last week. In each case, you will be asked to indicate how often you felt or thought a certain way. Although some of the questions are similar, there are differences between them and you should treat each one as a separate question. The best approach is to answer fairly quickly.

Questions:

1. In the last week, how often have you been upset because of something that happened unexpectedly?

2. In the last week, how often have you felt that you were unable to control the important things in your life?

3. In the last week, how often have you felt nervous and “stressed”?

4. In the last week, how often have you felt confident about your ability to handle your personal problems?

5. In the last week, how often have you felt that things were going your way?

6. In the last week, how often have you found that you could not cope with all the things that you had to do?
7. In the last week, how often have you been able to control irritations in your life?
8. In the last week, how often have you felt that you were on top of things?
9. In the last week, how often have you been angered because of things that were outside of your control?
10. In the last week, how often have you felt difficulties were piling up so high that you could not overcome them?

SCALE: never (0) - very often (4)
Coding: reverse scores for questions 4, 5, 7, and 8. On these 4 questions, change the scores as follows: 0 = 4, 1 = 3, 2 = 2, 3 = 1, 4 = 0

D. PHQ-8

Questions: Over the last two weeks, how often have you been bothered by any of the following problems?

1. Little interest or pleasure in doing things?
2. Feeling down, depressed, or hopeless?
3. Trouble falling or staying asleep, or sleeping too much?
4. Feeling tired or having little energy?
5. Poor appetite or overeating?
6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down?
7. Trouble concentrating on things, such as reading the newspaper or watching television?
8. Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?

SCALE: Not at all (0), Several days (1), More than half the days (3), Nearly every day (4)

E. GAD-7

Questions: Over the last week, how often have you been bothered by the following problems?

1. Feeling nervous, anxious, or on edge
2. Not being able to stop or control worrying
3. Worrying too much about different things
4. Trouble relaxing
5. Being so restless that it's hard to sit still
6. Becoming easily annoyed or irritable
7. Feeling afraid as if something awful might happen

SCALE: Not at all sure (0), Several days (1), Over half the days (2), Nearly every day (3)

F. WHO-5

Introduction: Please indicate for each of the five statements which is closest to how you have been feeling over the last week. Notice that higher numbers mean better well-being.

Questions:

1. I have felt cheerful and in good spirits
2. I have felt calm and relaxed
3. I have felt active and vigorous
4. I woke up feeling fresh and rested
5. my daily life has been filled with things that interest me

SCALE: All of the time (5), Most of the time (4), More than half of the time (3), Less than half of the time (2), Some of the time (1), At no time (0)

G. CD-RISC 10

Introduction: Please indicate how much you agree with the following statements as they apply to you over the last week. If a particular situation has not occurred recently, answer according to how you think you would have felt.

Questions:

1. I am able to adapt when changes occur.
2. I can deal with whatever comes my way.
3. I try to see the humorous side of things when I am faced with problems.
4. Having to cope with stress can make me stronger.
5. I tend to bounce back after illness, injury, or other hardships.
6. I believe I can achieve my goals, even if there are obstacles.
7. Under pressure, I stay focused and think clearly.
8. I am not easily discouraged by failure.
9. I think of myself as a strong person when dealing with life's challenges and difficulties.
10. I am able to handle unpleasant or painful feelings like sadness, fear, and anger.

SCALE: Not true at all (0), Rarely true (1), Sometimes true (2), Often true (3), True nearly all the time (4)

H. Additional scales

- BIG-5 Personality **[ONBOARDING ONLY]**
- MISS
    - Experience vs Evaluation scale

A7. Main results

A. Study 1

	Treatment				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	7.03	7.1	0.07	**0.42663**	273
Worthwhileness	7.26	7.39	0.13	**0.21763**	272
Happiness yesterday	6.95	7.17	0.22	**0.16173**	273
Anxiety yesterday	5.56	4.86	-0.7	**0.00426**	272
Happiness in general	7.22	7.43	0.21	**0.0727**	273
Anxiety in general	5.71	5.27	-0.44	**0.06072**	273

	Control				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	7.15	7.08	-0.07	**0.87286**	122
Worthwhileness	7.38	7.57	0.19	**0.37862**	122
Happiness yesterday	7.43	7.12	-0.31	**0.38112**	122
Anxiety yesterday	5.21	5.47	0.26	**0.42348**	122
Happiness in general	7.27	7.41	0.14	**0.41552**	122
Anxiety in general	5.22	5.48	0.26	**0.45328**	122

	Treatment effect		
Mann-Whitney	**Difference**	**P (two-sided)**	**Cohen's *d***
Life satisfaction	0.14	**0.25286**	0.09
Worthwhileness	-0.06	**0.92776**	0.03
Happiness yesterday	0.53	**0.0707**	0.26
Anxiety yesterday	-0.96	**0.00222**	0.36
Happiness in general	0.07	**0.3714**	0.04
Anxiety in general	-0.7	**0.01428**	0.30

B. Study 2

	Treatment				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	6.79	6.73	-0.06	**0.99534**	255
Worthwhileness	7.09	6.91	-0.18	**0.3007**	253
Happiness yesterday	6.58	6.71	0.13	**0.31294**	252
Anxiety yesterday	4.86	4.5	-0.36	**0.12146**	255
Happiness in general	6.78	6.82	0.04	**0.56686**	253
Anxiety in general	5.17	4.67	-0.5	**0.01656**	255

	Control				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	6.89	6.51	-0.38	**0.14538**	93
Worthwhileness	6.84	6.77	-0.07	**0.88188**	92
Happiness yesterday	6.6	6.19	-0.41	**0.25248**	93
Anxiety yesterday	5.18	5.08	-0.1	**0.79112**	93
Happiness in general	6.77	6.57	-0.2	**0.48244**	93
Anxiety in general	5.52	5.44	-0.08	**0.97056**	91

	Treatment effect		
Mann-Whitney	**Difference**	**P (two-sided)**	**Cohen's *d***
Life satisfaction	0.32	**0.1085**	0.20
Worthwhileness	-0.11	**0.82106**	0.06
Happiness yesterday	0.54	**0.0812**	0.27
Anxiety yesterday	-0.26	**0.6055**	0.10
Happiness in general	0.24	**0.17726**	0.15
Anxiety in general	-0.42	**0.037**	0.18

C. Study 3

	Treatment				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	6.5	7.1	0.6	**0.02064**	82
Worthwhileness	6.64	6.87	0.23	**0.51118**	83
Happiness yesterday	6.12	6.7	0.58	**0.02568**	83
Anxiety yesterday	5.71	5.45	-0.26	**0.59252**	84
Happiness in general	6.54	6.79	0.25	**0.36622**	84
Anxiety in general	5.73	5.67	-0.06	**0.86112**	84
GAD-7	8.66	6.79	-1.87	**0.02714**	82
PSS-10	21.59	19.3	-2.29	**0.02234**	80
PHQ-8	9.48	7.85	-1.63	**0.05314**	79
WHO-5	44.14	48.14	4	**0.31056**	84

	Control				
Mann-Whitney	**Mean onb**	**Mean exit**	**Diff**	**P (two-sided)**	**N**
Life satisfaction	6.5	6.78	0.28	**0.3113**	125
Worthwhileness	6.63	6.61	-0.02	**0.5878**	126
Happiness yesterday	5.94	6.5	0.56	**0.0694**	129
Anxiety yesterday	5.56	5.13	-0.43	**0.16266**	127
Happiness in general	6.19	6.39	0.2	**0.27734**	129
Anxiety in general	5.61	5.62	0.01	**0.92918**	128
GAD-7	8.35	8.07	-0.28	**0.62128**	124
PSS-10	20.75	19.01	-1.74	**0.07566**	119
PHQ-8	9.2	8.22	-0.98	**0.15628**	123
WHO-5	43.07	46.64	3.57	**0.17274**	129

	Treatment effect		
Mann-Whitney	**Difference**	**P (two-sided)**	**Cohen's *d***
Life satisfaction	0.32	**0.06182**	0.18

Worthwhileness	0.25	**0.17072**	0.13
Happiness yesterday	0.02	**0.70608**	0.01
Anxiety yesterday	0.17	**0.6055**	0.07
Happiness in general	0.05	**0.89944**	0.03
Anxiety in general	-0.07	**0.97872**	0.03
GAD-7	-1.59	**0.0115**	0.30
PSS-10	-0.55	**0.44608**	0.14
PHQ-8	-0.65	**0.16876**	0.12
WHO-5	0.43	**0.71482**	0.09

[1] Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231-1236.

[2] Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests–An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*, 95-102.